

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 06:08:13 ; Search time 370 Seconds
(without alignments)
4984.946 Million cell updates/sec

Title: US-10-647-089-93

Perfect score: 321

Sequence: 1 ttgacccacagcgactaa.....acgaacccgacagcaaacag 321

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 4300275 seqs, 2872944193 residues

Total number of hits satisfying chosen parameters: 8600550

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA:*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US05_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US06_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09A_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09B_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09C_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_NEW_PUB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10A_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10B_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10C_PUBCOMB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US10D_PUBCOMB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US10E_PUBCOMB.seq.*
18: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
19: /cgn2_6/ptodata/2/pubpna/US11_NEW_PUB.seq.*
20: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
21: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	100.0	321	9	US-09-894-844-93
2	321	100.0	321	16	US-10-388-902-93
3	321	100.0	321	16	US-10-647-089-93
c	44.8	14.0	1599	15	US-10-156-761-2997
5	44.8	14.0	9025608	15	US-10-156-761-1
6	44	13.7	909	18	US-10-425-115-73640
7	43.8	13.6	3480	15	US-10-369-493-39849
8	43.8	13.6	3480	15	US-10-369-493-39849
9	43.8	13.6	3501	15	US-10-369-493-39098
10	43	13.4	1193	16	US-10-425-114-24011
11	43	13.4	1308	18	US-10-425-115-49576
c	42.8	13.3	4200	9	US-09-815-242-4124

C	13	42.8	13.3	4200	16	US-10-282-122A-7392	Sequence 7392, Ap
	14	42.8	13.3	10232	16	US-10-282-122A-25487	Sequence 25487, A
	15	42	13.1	1959	15	US-10-156-761-4513	Sequence 4513, Ap
C	16	42	13.1	9025608	15	US-10-156-761-1	Sequence 1, Appli
	17	41.6	13.0	996	16	US-10-282-122A-11268	Sequence 11268, A
C	18	41.2	12.8	936	17	US-10-437-963-76527	Sequence 76527, A
	19	41	12.8	400	9	US-09-867-701-4968	Sequence 4968, Ap
	20	40.8	12.7	1232	17	US-10-437-963-29422	Sequence 29422, A
	21	40.6	12.6	2580	15	US-10-156-761-3895	Sequence 3895, Ap
C	22	40.6	12.6	3044	17	US-10-437-963-79637	Sequence 79637, A
	23	40.4	12.6	951	16	US-10-282-122A-26156	Sequence 26156, A
	24	40.4	12.6	954	9	US-09-712-363-39	Sequence 39, Appli
	25	40.4	12.6	954	16	US-10-282-122A-28300	Sequence 28300, A
	26	40.4	12.6	45191	15	US-10-080-170-649	Sequence 649, App
	27	40.4	12.6	45191	17	US-10-080-170-649	Sequence 649, App
	28	40.4	12.5	45191	18	US-10-468-356-649	Sequence 649, App
C	29	40.2	12.5	307	18	US-10-425-115-131105	Sequence 131105, A
	30	40.2	12.5	1803	15	US-10-156-761-4657	Sequence 4657, Ap
	31	40.2	12.5	1824	17	US-10-767-701-13070	Sequence 13070, A
C	32	39.8	12.4	1300	17	US-10-437-963-32867	Sequence 22867, A
	33	39.6	12.3	1318	16	US-10-425-114-14884	Sequence 14884, A
C	34	39.6	12.3	1446	15	US-10-156-761-4324	Sequence 4324, Ap
	35	39.6	12.3	1460	16	US-10-425-114-14893	Sequence 14893, A
C	36	39.6	12.3	1512	13	US-10-124-800-31	Sequence 31, Appli
	37	39.6	12.3	1512	18	US-10-810-352-31	Sequence 31, Appli
C	38	39.6	12.3	2073	16	US-10-425-114-29125	Sequence 29125, A
	39	39.6	12.3	4509	13	US-10-124-800-5	Sequence 5, Appli
C	40	39.6	12.3	4509	18	US-10-810-352-5	Sequence 5, Appli
	41	39.6	12.3	4512	13	US-10-124-800-27	Sequence 27, Appli
C	42	39.6	12.3	4512	18	US-10-810-352-27	Sequence 27, Appli
	43	39.6	12.3	4767	15	US-10-331-061-76	Sequence 76, Appli
C	44	39.4	12.3	1488	9	US-09-815-242-7862	Sequence 7862, Ap
	45	39.4	12.3	1488	16	US-10-282-122A-30375	Sequence 30375, A

ALIGNMENTS

RESULT 1
US-09-894-844-93
; Sequence 93, Application US/09894844
; Patent No. US20020176873A1
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of the M. Tuberculosis Complex
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-09-894-844-93

Query Match 100.0%; Score 321; DB 9; Length 321;
Best Local Similarity 100.0%; Pred. No. 9.3e-90;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGACCCACAGCGCCTAAACGCCAGCGCAGCATCGCCGAGGCTCAAGCCCGCGGT 60

Db 1 TTGACCCACAGCGCCTAAACGCCAGCGCAGCATCGCCGAGGCTCAAGCCCGCGGT 60

QY 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGCGACGTTTCGTCGCGGAGCAGCGC 120

; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 2997
 ; LENGTH: 1599
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: (1)..(1599)
 US-10-156-761-2997

Query Match 14.0%; Score 44.8; DB 15; Length 1599;
 Best Local Similarity 51.5%; Pred. No. 0.00051;
 Matches 103; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
 QY 42 AGGGCTCAAGCCCGCGGTGCGAATCGGTTGGCGGCAACATGTTGGCCGGCCGACGT 101
 DB 415 AGGGCAGCCCCACCGCGCGCCACCGCGTCTCGCGACCGCGTCCGGCGCGCGACCA 356
 QY 102 TCCGTCGCGGAGCGCGCGCCCAACCGGCGAGCGGACCTCGAGGCTATCCGCGCGAGC 161
 DB 355 CCGCGCGGAGCACCAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 296
 QY 162 GTACGCGGAGATGTTGGCGACATCACGAAATCGACGACGACGACGCGGAACTGGCGCT 221
 DB 295 AGCCACGACGAGGTCGCGAGAGAGAGCTAGCCGATGACGACGACGCGCCACCGTCCGA 236
 QY 222 GTTGTGCGATCTCGACG 241
 DB 235 GTTGGCGCGCAGATCGAGG 216

RESULT 5
 US-10-156-761-1
 ; Sequence 1, Application US/10156761
 ; Publication No. US20030119018A1
 ; GENERAL INFORMATION:
 ; APPLICANT: OMURA, SATOSHI
 ; APPLICANT: IKEDA, HARUO
 ; APPLICANT: ISHIKAWA, JUN
 ; APPLICANT: HORIKAWA, HIROSHI
 ; APPLICANT: SHIBA, TADAYOSHI
 ; APPLICANT: SAKAKI, YOSHIYUKI
 ; APPLICANT: HATTORI, MASAHIRA
 ; TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
 ; FILE REFERENCE: 249-262
 ; CURRENT APPLICATION NUMBER: US/10/156,761
 ; CURRENT FILING DATE: 2002-05-29
 ; PRIOR APPLICATION NUMBER: JP 2001-204089
 ; PRIOR FILING DATE: 2001-05-30
 ; PRIOR APPLICATION NUMBER: JP 2001-272697
 ; PRIOR FILING DATE: 2001-08-02
 ; NUMBER OF SEQ ID NOS: 15109
 ; SEQ ID NO 1
 ; LENGTH: 9025608
 ; TYPE: DNA
 ; ORGANISM: Streptomyces avermitilis
 ; FEATURE:
 ; NAME/KEY: misc feature
 ; LOCATION: (4187715)
 ; OTHER INFORMATION: a, t, c, g, other or unknown
 US-10-156-761-1

Query Match 14.0%; Score 44.8; DB 15; Length 9025608;
 Best Local Similarity 51.5%; Pred. No. 0.00078;
 Matches 103; Conservative 0; Mismatches 97; Indels 0; Gaps 0;

QY 42 AGGGCTCAAGCCCGCGGTGCGAATCGGTTGGCGGCAACATGTTGGCGCGCGACGT 101
 DB 3766916 AGGGCAGCCCGCCACCGCGCGCCACCGCGTCTCGCGGCGGCGCGACCA 3766975
 QY 102 TCCGTCGCGGAGCGCGCGCCCAACCGGCGAGCGGACCTCGAGGCTATCCGCGGAGC 161
 DB 3766976 CCGCGCGGAGCACCAGCGCGCGCGGCGGCGGCGGCGGCGGCGGCGGCGG 3767035
 QY 162 GTACGCGGAGATGTTGGCGACATCACAGAAATCGACGACGACACAGCCGAACTGGCGCT 221
 DB 3767036 AGCCACGACGAGGTCGCGAGAGGAGCTAGCCGATGGACGACGCGCCACCGTCCGA 3767095
 QY 222 GTTGTGCGATCTCGACG 241
 DB 3767096 GTTGGCGCGCAGATCGAGG 3767115

RESULT 6
 US-10-425-115-73640
 ; Sequence 73640, Application US/10425115
 ; Publication No. US20040214272A1
 ; GENERAL INFORMATION:
 ; APPLICANT: La Rosa, Thomas J.
 ; APPLICANT: Kowalic, David K.
 ; APPLICANT: Zhou, Yihua
 ; APPLICANT: Cao, Yongwei
 ; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
 ; TITLE OF INVENTION: Plants
 ; FILE REFERENCE: 38-21(53222)B
 ; CURRENT APPLICATION NUMBER: US/10/425,115
 ; CURRENT FILING DATE: 2003-04-28
 ; NUMBER OF SEQ ID NOS: 369326
 ; SEQ ID NO 73640
 ; LENGTH: 909
 ; TYPE: DNA
 ; ORGANISM: Zea mays
 ; FEATURE:
 ; OTHER INFORMATION: Clone ID: MRT4577_167159C.1
 US-10-425-115-73640

Query Match 13.7%; Score 44; DB 18; Length 909;
 Best Local Similarity 54.3%; Pred. No. 0.00087;
 Matches 89; Conservative 0; Mismatches 75; Indels 0; Gaps 0;
 QY 96 CGAGCTTCGTCGCGCGGAGCGCGCGCCCAACCGGCGAGCGGCGGCGGCGGCGGCGG 155
 DB 578 CCGGATTTGGGCGGCGGCGGCGGCTGCGGTACGCGGAGCGGTCGCGGAGCGCGGCGG 637
 QY 156 CCGAGCGTACGCGCGGAGATGTTGGCGACATCATCAAGAAATCGACGACGACGACGCGGAACT 215
 DB 638 GTTGGAGTGGCGGCGGCTGATGTTGGCGCAGCTGCACGAGCTCAAGGGGGGACCTCGCGGAGC 697
 QY 216 GCGCGTGTGTCGATGTCATCTCGACGATGAGCAGCGCGGCTTG 259
 DB 698 TGAGCGTCACTTCAGGAGGCTCCTGGCAGAGAGCGCCCTCTCTCG 741

RESULT 7
 US-10-369-493-39466
 ; Sequence 39466, Application US/10369493
 ; Publication No. US20030233675A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Cao, Yongwei
 ; APPLICANT: Hinkle, Gregory J.
 ; APPLICANT: Slater, Steven C.
 ; APPLICANT: Goldman, Barry S.
 ; APPLICANT: Chen, Xianfeng
 ; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
 ; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
 ; FILE REFERENCE: 38-10(52052)B
 ; CURRENT APPLICATION NUMBER: US/10/369,493
 ; CURRENT FILING DATE: 2003-02-28

Db 451 CGCGGACCTCGTGGCCCTTGGCGGCGCCCGGTTCCCGTCCGCTGGCGCGCGCG 510
Qy 287 CGTATCACTCCCGACGAAACCCGACACAA 317
Db 511 ACTCCCGCGCTCCGACGCGCGGACGTCGA 541

RESULT 11

US-10-425-115-49576
; Sequence 49576, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 49576
; LENGTH: 1308
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MXT4577_145214C.1
US-10-425-115-49576

Query Match 13.4%; Score 43; DB 18; Length 1308;
Best Local Similarity 50.2%; Pred. No. 0.0018;
Matches 106; Conservative 0; Mismatches 105; Indels 0; Gaps 0;
Qy 107 CGCGGACGCGCGCGCCCAACGGCAGCGGACCTCGAGGCTATCCGCCGAGGTCAG 166
Db 331 CGCGGAGATCAACCTGCTCCCTCGCCCGGGGAGCGCTTCGACGCGCTGGCCCGGCCAAG 390
Qy 167 CGGAGATGTCGGGACATCACAGAAATCGACGACACAGCGAACTGGCGCTGTGT 226
Db 391 CGGCGCTGGAGCGCGGTGTCGCCGGCGCTGCTCTCGCCGACGCGCTGGCGCTAGCGG 450
Qy 227 CGATCATCTCGACATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATC 286
Db 451 CGCGGACCTCGTGGCGCCCTTGGCGGCGCCCGGTTCCCGTCCGCTGGCGCGCGCG 510
Qy 287 CGTATCACTCCCGACGAAACCCGACACAA 317
Db 511 ACTCCCGCGCTCCGACGCGCGGACGTCGA 541

RESULT 12

US-09-815-242-4124/c
; Sequence 4124, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; TITLE OF INVENTION: Prokaryotes
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848

; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: Fast-Seq for Windows Version 4.0
; SEQ ID NO 4124
; LENGTH: 4200
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-4124

Query Match 13.3%; Score 42.8; DB 9; Length 4200;
Best Local Similarity 50.5%; Pred. No. 0.0022;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
Qy 48 CAACGCCCCGCTCGGAATCGCGTTGGCGGCAACATGGTTGGCGCGCGCGCTTCCGTC 107
Db 1519 CCACGCTCTCGCAGCGAAACGATGATCGGCTCGCCGTTGGCGGCGCACAGGATGTTGTTGG 1460
Qy 108 CGCGGACGCGCGCGCCCAACGGCAGCGGACCTCGAGGCTATCCGCCGAGGTCAGCG 167
Db 1459 TGGACATCATCAGCGCGCGCGCTTCCAGCTGGGCTCGAGGTCAGCGGAACTGGAGCG 1400
Qy 168 CGAGATGTCGGCGACATCACAGAAATCGACGACACACAGCGAACTGGCGCTGTTGTC 227
Db 1399 CCATCTGTCACCGTCGAAGTCGCGGTTGTACGCGGCGCACACCGCGGTCAGCTGGA 1340
Qy 228 GATCATCTCGACCATGAGCGGCC 253
Db 1339 TCGCCTTCCCTTCGATGAGGACCGCG 1314

RESULT 13

US-10-282-122A-7392/c
; Sequence 7392, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23

```

; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7392
; LENGTH: 4200
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-10-282-122A-7392

Query Match          13.3%   Score 42.8;   DB 16;   Length 4200;
Best Local Similarity 50.5%;   Pred. No. 0.0022;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

Qy 48 CAACGCCCCCGCTCGGAATCGGTTGGCGGCGCAACATGTTGGCGCGCCGACGTTCCGTC 107
Db 1519 CCAGTCTCTGCGACGGAAACGATGATCGGCTCGCCGTTGGCGGCGACAGATGTTGTTGG 1460

Qy 108 CGCGGACGAGCGCGCGCCCAACGGCAGCGGACCTTCAGGCTATCCGCGAGCGTACGC 167
Db 1459 TGGACATCATCAGCGCGCGCGCTTCCAGCTGGGGCTTCGAGGGTCAGCGGAACCGTGGACGG 1400

Qy 168 CGAGATGGTGGCGACATCATACGAAATCGACGACACAGCCGAACTGGCGCTGTTGTC 227
Db 1399 CCATCTGGTCACCGTCAAGTTCGGCGTTGTACCGGGGCGACAGACGGGTGAGCTGGA 1340

Qy 228 GATGCAATCTCGACGATGAGCAGCGCC 253

Db 1339 TCGCCTTGCCTTCGATGAGGACCGGC 1314

```

RESULT 14

US-10-282-122A-25487

Sequence 25487, Application US/10282122A

Publication No. US20040029129A1

GENERAL INFORMATION:

APPLICANT: Wang, Liangsu

APPLICANT: Zamudio, Carlos

APPLICANT: Malone, Cheryl

APPLICANT: Haselbeck, Robert

APPLICANT: Ohlsen, Kari

APPLICANT: Zyskind, Judith

APPLICANT: Wall, Daniel

APPLICANT: Trawick, John

APPLICANT: Carr, Grant

APPLICANT: Yamamoto, Robert

APPLICANT: Forsyth, R.

APPLICANT: Xu, H.

FILE OF INVENTION: Identification of Essential Genes in Microorganisms

TITLE REFERENCE: ELITRA.034A

CURRENT APPLICATION NUMBER: US/10/282,122A

CURRENT FILING DATE: 2003-02-20

PRIOR APPLICATION NUMBER: 60/191,078

PRIOR FILING DATE: 2000-03-21

PRIOR APPLICATION NUMBER: 60/206,848

PRIOR FILING DATE: 2000-05-23

PRIOR APPLICATION NUMBER: 60/207,727

PRIOR FILING DATE: 2000-05-26

PRIOR APPLICATION NUMBER: 60/230,335

PRIOR FILING DATE: 2000-09-06

PRIOR APPLICATION NUMBER: 60/230,347

PRIOR FILING DATE: 2000-09-09

PRIOR APPLICATION NUMBER: 60/242,578

PRIOR FILING DATE: 2000-10-23

PRIOR APPLICATION NUMBER: 60/253,625

PRIOR FILING DATE: 2000-11-27

```

; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25487
; LENGTH: 10232
; TYPE: DNA
; ORGANISM: Mycobacterium avium
US-10-282-122A-25487

Query Match      13.3%; Score 42.8; DB 16; Length 10232;
Best Local Similarity 47.7%; Pred. No. 0.0023;
Matches 125; Conservative 0; Mismatches 137; Indels 0; Gaps 0;

Qy      20  AACGCCAGCAGCAGCATCCGCGCAGGGGCTCAACGCCCGCGGTTCGAATCGGATTGGGGCGC 79
Db      1271  AACTGTGTACCGCGCAGCAGCCCCGCGTGTATCACCGGTGAGACAACCGTCCGCGCGCTGA 1330

Qy      80  AACATGTTGGCCGGCGCAGCTTCGTTCGCCGAGCAGCGCCGCCGCCAACGCGCAGCGCG 139
Db      1331  CCGAGCCCGCACCGCCCGGTGTGATCCCCCAGGCCCTTCGCCGGAACAGTTCGAGGCCA 1390

Qy      140  ACCTCGAGGCTATCCCGCGAGCGCTACCGCGAGATGTGTGGCGACATCACAGAAATCGACG 199
Db      1391  CCGCCCGCAGCGGTGTGGCGGTGACGTTCCGGGCGCACCTCGCTGACCTAGCGCCAGCTCGACG 1450

Qy      200  ACGACACAGCCGCAACTGGCGCTCTGTTCGATGCATCTCGACGATGAGCAGCGCGCGCTTG 259
Db      1451  AGCGGTTCACACCGCTTGGGCCATCTGCTGCCGCCACACGGCGTGGGCCCGGGCGACTGCG 1510

Qy      260  AGCGCGGGATGAAGCTCGGCTG 281
Db      1511  TGGCGGTGATGTTCCCGCGCTG 1532

```

```

RESULT 15
US-10-156-761-4513
, Sequence 4513, Application US/10156761
, Publication No. US20030119018A1
, GENERAL INFORMATION:
, APPLICANT: OMURA, SATOSHI
, APPLICANT: IKEDA, HARUO
, APPLICANT: ISHIKAWA, JUN
, APPLICANT: HORIKAWA, HIROSHI
, APPLICANT: SHIBA, TADAYOSHI
, APPLICANT: SAKAKI, YOSHIYUKI
, APPLICANT: HAYTORI, MASASHIRA
, TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
, FILE REFERENCE: 249-262
, CURRENT APPLICATION NUMBER: US/10/156,761
, CURRENT FILING DATE: 2002-05-29
, PRIOR APPLICATION NUMBER: JP 2001-204089
, PRIOR FILING DATE: 2001-05-30
, PRIOR APPLICATION NUMBER: JP 2001-272697
, PRIOR FILING DATE: 2001-08-02
, NUMBER OF SEQ ID NOS: 15109
, SEQ ID NO 4513
, LENGTH: 1959
, TYPE: DNA
, ORGANISM: Streptomyces avermitilis
, FEATURE:
, NAME/KEY: CDS
, LOCATION: (1)..(1959)
US-10-156-761-4513

```

Query Match 13.1%; Score 42; DB 15; Length 1959;
Best Local Similarity 50.5%; Pred. No. 0.0038;
Matches 102; Conservative 0; Mismatches 100; Indels 0; Gaps 0;

```

Qy 20 AACGCCAGCCAGCCATCGCCGAGGGCTCAACGCCCCCGCGTCGGAATCGCGTTGGCGGC 79
Db 365 AGCAGGACACCTCAACCGTCAACCGACCTGGGGGACGCCCGTCGGCGCGCTTCGGGGGCC 424
Qy 80 AACATGGTTGGCCGCCGACGTTCCGTCGGCCGAGCAGGCGCGGCCCAACGGCAGCGCG 139
Db 425 GGCACGGACCCGACCCCGACCTTGGCGCCCTTCCCGCGCCACGCCCGCCCGCCCGCG 484
Qy 140 ACCTCGAGCTATCCGCCGAGCGTACGCCGAGATGGTGGCGACATCACACGAAATCGACG 199
Db 485 TCGAGGACGACCCCGGAGCATGACACCCGCGCATCTGGCGCCCGGACCCAGTTCGACC 544
Qy 200 ACGACACAGCCGAACTGGCGCT 221
Db 545 CGTGGTTCGCGGACCTGGTGGT 566

```

Search completed: February 10, 2005, 07:36:14
Job time : 390 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 05:25:02 ; Search time 75 seconds
(without alignments)
3042.176 Million cell updates/sec

Title: US-10-647-089-93

Perfect score: 321

Sequence: 1 ttgaccacagcgcaactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 824507 seqs, 355394441 residues

Total number of hits satisfying chosen parameters: 1649014

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- Issued Patents NA:*
- 1: /cgn2_6/ptodata/1/ina/5A_COMB.seq:*
 - 2: /cgn2_6/ptodata/1/ina/5B_COMB.seq:*
 - 3: /cgn2_6/ptodata/1/ina/6A_COMB.seq:*
 - 4: /cgn2_6/ptodata/1/ina/6B_COMB.seq:*
 - 5: /cgn2_6/ptodata/1/ina/PCTUS_COMB.seq:*
 - 6: /cgn2_6/ptodata/1/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	100.0	321	4	US-09-894-844-93
C 2	321	100.0	4403765	3	Sequence 93, Appl
C 3	321	100.0	4411529	3	Sequence 2, Appli
C 4	42.8	13.3	4212	4	US-09-103-840A-1
C 5	42.8	13.3	4242	4	US-09-252-991A-4748
C 6	40.4	12.6	4403765	3	Sequence 4697, Ap
C 7	40.4	12.6	4411529	3	Sequence 2, Appli
C 8	40.2	12.5	732	4	US-09-103-840A-1
C 9	40.2	12.5	945	4	US-09-252-991A-6298
C 10	40.2	12.5	1602	4	US-09-252-991A-6167
C 11	39.8	12.4	594	4	US-09-252-991A-6382
C 12	39.8	12.4	966	4	US-09-252-991A-12043
C 13	39.8	12.4	1179	4	US-09-252-991A-11896
C 14	39.6	12.3	4767	4	US-09-252-991A-12017
C 15	39.4	12.3	1470	4	US-09-231-899-76
C 16	39.4	12.3	1473	4	US-09-252-991A-6325
C 17	39.4	12.3	1512	4	US-09-252-991A-6144
C 18	39.2	12.2	1320	4	US-09-724-797-85
C 19	38.8	12.1	408	4	US-09-252-991A-655
C 20	38.8	12.1	801	4	US-09-252-991A-698
C 21	38.8	12.1	1356	4	US-09-252-991A-630
C 22	38.8	12.1	2955	4	US-09-252-991A-684
C 23	38.8	12.1	3150	4	US-09-252-991A-727
C 24	38.4	12.0	1371	4	US-09-252-991A-13954
C 25	38.4	12.0	2181	4	US-09-252-991A-13742
C 26	38.4	12.0	2277	4	US-09-252-991A-13403
C 27	38.4	12.0	2549	4	US-09-638-203-1

C 28	38.4	12.0	2570	4	US-10-140-002-539	Sequence 539, App
C 29	38.2	11.9	888	4	US-09-252-991A-16512	Sequence 16512, A
C 30	38.2	11.9	1512	4	US-09-252-991A-16092	Sequence 16092, A
C 31	38.2	11.9	2400	4	US-09-252-991A-15991	Sequence 15991, A
C 32	37.8	11.8	4257	2	US-08-690-473-1	Sequence 1, Appli
C 33	37.8	11.8	4257	3	US-09-259-821A-1	Sequence 1, Appli
C 34	37.8	11.8	4257	3	US-08-843-659-1	Sequence 1, Appli
C 35	37.8	11.8	4257	4	US-08-825-288A-1	Sequence 11, Appl
C 36	37.8	11.6	12001	1	US-08-458-568A-11	Sequence 943, App
C 37	37.2	11.5	1488	4	US-09-489-039A-943	Sequence 895, App
C 38	37	11.5	1470	4	US-09-252-991A-896	Sequence 1137, Ap
C 39	37	11.5	1662	4	US-09-252-991A-1137	Sequence 23, Appl
C 40	37	11.5	1693	3	US-09-320-878-23	Sequence 19, Appl
C 41	37	11.5	1693	4	US-09-141-908-19	Sequence 23, Appl
C 42	37	11.5	1693	4	US-09-657-440-23	Sequence 945, App
C 43	37	11.5	1839	4	US-09-252-991A-945	Sequence 7, Appli
C 44	37	11.5	44377	2	US-08-804-227C-7	Sequence 1, Appli
C 45	37	11.5	44377	2	US-08-804-198-1	

ALIGNMENTS

RESULT 1
US-09-894-844-93
; Sequence 93, Application US/09894844
; Patent No. 6686166
; GENERAL INFORMATION:
; APPLICANT: Behr, Marcel
; APPLICANT: Small, Peter
; APPLICANT: Schoolnik, Gary
; APPLICANT: Wilson, Michael A.
; TITLE OF INVENTION: Molecular Differences Between Species of
; FILE REFERENCE: STAN102CON
; CURRENT APPLICATION NUMBER: US/09/894,844
; CURRENT FILING DATE: 2001-06-27
; PRIOR APPLICATION NUMBER: 09/318,191
; PRIOR FILING DATE: 1999-05-25
; PRIOR APPLICATION NUMBER: 60/097,936
; PRIOR FILING DATE: 1998-08-25
; NUMBER OF SEQ ID NOS: 137
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 93
; LENGTH: 321
; TYPE: DNA
; ORGANISM: Mycobacteria tuberculosis
US-09-894-844-93

Query Match	100.0%	Score 321;	DB 4;	Length 321;
Best Local Similarity	100.0%	Pred. No. 3.6e-75;		
Matches 321;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
Qy	1	TTGACCCAAACGCGCACTAAACGCGGAGGCTACGCGGAGGCTCAACGCCCGCGT	60	
Db	1	TTGACCCAAACGCGCACTAAACGCGGAGGCTACGCGGAGGCTCAACGCCCGCGT	60	
Qy	61	CGGAATCGCGTTGGCGGCAACATGTTGGCGGCGGAGCTTCGCTCCGCGGAGCAGCGC	120	
Db	61	CGGAATCGCGTTGGCGGCAACATGTTGGCGGCGGAGCTTCGCTCCGCGGAGCAGCGC	120	
Qy	121	CGCGCCAAACGCGGCGGAGCTTCGAGGCTATCCGCGGAGGCTACGCCGAGATGGTGGC	180	
Db	121	CGCGCCAAACGCGGCGGAGCTTCGAGGCTATCCGCGGAGGCTACGCCGAGATGGTGGC	180	
Qy	181	ACATCACAAGAAATCGACGACGACACACGCGGAGTGGCGGCTGTTGTCATGATCTCGAC	240	
Db	181	ACATCACAAGAAATCGACGACGACACACGCGGAGTGGCGGCTGTTGTCATGATCTCGAC	240	
Qy	241	GATGAGCAGCGCGCGCTTTGAGCGGGATGAAGCTCGGCTCGCATCCGTCATCATTCCCC	300	
Db	241	GATGAGCAGCGCGCGCTTTGAGCGGGATGAAGCTCGGCTCGCATCCGTCATCATTCCCC	300	

QY 306 ACCGACAGCAACAG 321
Db 1129665 GCCGGCAGCAAGAG 1129680

RESULT 8

US-09-252-991A-6298/c
; Sequence 6298, Application US/09252991A
; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6298

; LENGTH: 732

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6298

Query Match 12.5%; Score 40.2; DB 4; Length 732;

Best Local Similarity 49.3%; Pred. No. 0.1;

Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 90 GCCGGCCGACGTTCCGTCGCCGAGCAGCGCGCCCAACGCGGACCTCGAGGC 149

Db 692 GCGGATCGCGTTGGCAGCGCCACACAGCGGCTCGCGCCAGGCCACCTCGA 633

QY 150 TATCCGCGGAGCTAGCCGAGATGTTGGCGACATCACAGAAATCGACGACACAGC 209

Db 632 TGCGGACCTGGCTTCGCGCGGCGGAGCGGCCAGCCACATCGATCGCGCTTCGG 573

QY 210 CGAATCGCGCTGTTGTCGATGATCTCGAGATGAGCAGCGCGGCTTGAGCGGGGAT 269

Db 572 CGATCTGCGCCCATTCGCGCGGAGCGCTTCGAGCAGCTCCAGTTGCACCTCCGGAT 513

QY 270 GAAGCTCGGCTGGCATCCGTCATCACTTCCCGCA 302

Db 512 GCGCTCGGGAACAGTCGATCGCCTCGGTGA 480

RESULT 9

US-09-252-991A-6167

; Sequence 6167, Application US/09252991A

; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6167

; LENGTH: 945

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6167

Query Match 12.5%; Score 40.2; DB 4; Length 945;

Best Local Similarity 49.3%; Pred. No. 0.11;

Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 90 GCCGGCCGACGTTCCGTCGCCGAGCAGCGCGCCCAACGCGGACCTCGAGGC 149
Db 420 GCGGATCGCGTTGGGACAGCGCCACACAGCGGCTCGCGGCCAGGCCACCTCGA 479
QY 150 TATCCGCGGAGCTAGCCGAGATGTTGGCGACATCACAGAAATCGACGACACAGC 209
Db 480 TCGGACCTGGGCTTCGCGCGGCGGAGCGGCCCGCCACATCGATCGCGCTTCGG 539
QY 210 CGAATCGGCGCTGTTGTCGATGATCTCGACGATGAGCAGCGCGGCTTGAGCGGGGAT 269
Db 540 CGATCTGCGCCCATTCGCGCGGCGGAGCGCTTCGAGCAGCTCCAGTTGCACCTCCGGAT 599
QY 270 GAAGCTCGGCTGGCATCCGTCATCACTTCCCGCA 302
Db 600 GCGCTCGGGAACAGTCGATCGCCTCGGTGA 632

RESULT 10

US-09-252-991A-6382/c

; Sequence 6382, Application US/09252991A

; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ ID NOS: 33142

; SEQ ID NO 6382

; LENGTH: 1602

; TYPE: DNA

; ORGANISM: Pseudomonas aeruginosa

US-09-252-991A-6382

Query Match 12.5%; Score 40.2; DB 4; Length 1602;

Best Local Similarity 49.3%; Pred. No. 0.12;

Matches 105; Conservative 0; Mismatches 108; Indels 0; Gaps 0;

QY 90 GCCGGCCGACGTTCCGTCGCCGAGCAGCGCGCCCAACGCGGAGCTCGAGGC 149

Db 1222 GCGGATCGCGTTGGGACAGCGGCCACACAGCGGCTCGCGGCCAGGCCACCTCGA 1163

QY 150 TATCCGCGGAGCTAGCCGAGATGTTGGCGACATCACAGAAATCGACGACACAGC 209

Db 1162 TCGGACCTGGGCTTCGCGCGGCGGAGCGGCCCGCCACATCGATCGCGCTTCGG 1103

QY 210 CGAATCGGCGCTGTTGTCGATGATCTCGACGATGAGCAGCGCGGCTTGAGCGGGGAT 269

Db 1102 CGATCTGCGCCCATTCGCGCGGCGGAGCGCTTCGAGCAGCTCCAGTTGCACCTCCGGAT 1043

QY 270 GAAGCTCGGCTGGCATCCGTCATCACTTCCCGCA 302

Db 1042 GCGCTCGGGAACAGTCGATCGCCTCGGTGA 1010

RESULT 11

US-09-252-991A-12043

; Sequence 12043, Application US/09252991A

; Patent No. 6551795

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

```

; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12043
; LENGTH: 594
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12043

Query Match
Best Local Similarity 12.4%; Score 39.8; DB 4; Length 594;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 AGCCAGCCATCCCGCAGGCTCAACGCCGCCGTCGGAATCGCGTTGGGGGCAACATG 85
DB 265 AGGAAGCTTCGCTGGGGCCGGTGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 324
QY 86 GTTGGCCGGCGACGTTCCGTCGCGAGCAGCGCGGCCCAACAGCGAGCGGACCTCG 145
DB 325 ACTTCCGAGCTCTCAGAGCCGACGCGGATGAAGGCGTCCGGAACGGCAGCGCCTGGTGC 384
QY 146 AGGCTATCCGCGAGCGTACGCCGAGATGGTGGCG 180
DB 385 AGGGTTTTCGCGCCTGTTCTCGCTGAGGTGGCG 419

RESULT 12
US-09-252-991A-11896/c
; Sequence 11896, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 11896
; LENGTH: 966
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-11896

Query Match
Best Local Similarity 12.4%; Score 39.8; DB 4; Length 966;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 AGCCAGCCATCCCGCAGGCTCAACGCCGCCGTCGGAATCGCGTTGGGGGCAACATG 85
DB 584 AGGAAGCTTCGCTGGGGCCGGTGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 525
QY 86 GTTGGCCGGCGACGTTCCGTCGCGAGCAGCGCGGCCCAACAGCGAGCGGACCTCG 145
DB 524 ACTTCCGAGCTCTCAGAGCCGACGCGGATGAAGGCGTCCGGAACGGCAGCGCCTGGTGC 465
QY 146 AGGCTATCCGCGAGCGTACGCCGAGATGGTGGCG 180
DB 464 AGGGTTTTCGCGCCTGTTCTCGCTGAGGTGGCG 430

RESULT 13
US-09-252-991A-12017
; Sequence 12017, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS

```

```

; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; PRIOR FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 12017
; LENGTH: 1179
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-12017

Query Match
Best Local Similarity 12.4%; Score 39.8; DB 4; Length 1179;
Matches 83; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 AGCCAGCCATCCCGCAGGCTCAACGCCGCCGTCGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 541
DB 482 AGGAAGCTTCGCTGGGGCCGGTGAAGACCCGCTGGAATTCCTGGCGGATGCCCT 541
QY 86 GTTGGCCGGCGACGTTCCGTCGCGAGCAGCGCGGCCCAACAGCGAGCGGACCTCG 145
DB 542 ACTTCCGAGCTCTCAGAGCCGACGCGGATGAAGGCGTCCGGAACGGCAGCGCCTGGTGC 601
QY 146 AGGCTATCCGCGAGCGTACGCCGAGATGGTGGCG 180
DB 602 AGGGTTTTCGCGCCTGTTCTCGCTGAGGTGGCG 636

RESULT 14
US-09-231-899-76/c
; Sequence 76, Application US/09231899
; Patent No. 6566583
; GENERAL INFORMATION:
; APPLICANT: Lassen, Michael
; APPLICANT: Metz, James G
; APPLICANT: Facciotti, Daniel
; TITLE OF INVENTION: SCHIZOCHYTRIUM PKS GENES
; FILE REFERENCE: CGNE.131.0205
; CURRENT APPLICATION NUMBER: US/09/231,899
; CURRENT FILING DATE: 1999-01-14
; EARLIER APPLICATION NUMBER: 60/048,650
; EARLIER FILING DATE: 1997-06-04
; EARLIER APPLICATION NUMBER: 09/090,793
; EARLIER FILING DATE: 1998-06-04
; NUMBER OF SEQ ID NOS: 86
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 76
; LENGTH: 4767
; TYPE: DNA
; ORGANISM: Schizochytrium aggregatum
US-09-231-899-76

Query Match
Best Local Similarity 12.3%; Score 39.6; DB 4; Length 4767;
Matches 117; Conservative 0; Mismatches 129; Indels 0; Gaps 0;

QY 70 GTTGGGGGCAACATGGTTGGCCGCCGACGTTCCGTCGCGAGCAGCGCGGCCCA 129
DB 4026 GGTGCGGACTGCTTGGCGACTGTTGACAGTGCCTGGTGACGATGAAGGCGGCCCAT 3967
QY 130 CGGAGCGCGACCTCGGAGGCTATCGCGGAGCGTACCGGAGATGGTGGCAGATCAC 189
DB 3966 GGTGAGCGCGCGCGCGCGCTCGCGGACCGACCGCGCGGCCCAACCGGAC 3907
QY 190 GAATTCGACGACGACACACGCGAACTGGCGCTGTTGTCGATGTCATCTCGAGATGAG 249
DB 3906 CGGAGGTGCGCGGGGTAGCGCACTCGCGGTGAGGCGGTTGGAGGTTGATGATG 3847
QY 250 CGCGGCTTGAGGGGGGATGAAGCTCGGCTGGCATCGGTATCACTTCCCCGACGAACC 309

```

```

Db      3846  CGGAGGATGACGTGGATGGGGGTTGTGGTGTGGCCGCCGAGTCAGCCTCGACAGC 3787
Qy      310  GACAGC 315
Db      3786  GATATC 3781

```

```

RESULT 15
US-09-252-991A-6325
; Sequence 6325, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Maic J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 6325
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-6325

```

```

Query Match      12.3%; Score 39.4; DB 4; Length 1470;
Best Local Similarity 54.5%; Pred.No. 0.19;
Matches 79; Conservative 0; Mismatches 66; Indels 0; Gaps 0;

Qy      1  TTGACCCACAGCGCCTAAGCCAGCCAGCCATCCCGCAGGGGCTCAAGCCCGCGT 60
Db      979  TTTCACCGGAGATCGCCGAGGGCCAGCGGCATCGCGCCCGCAGGCCCTTGAGGGT 1038

Qy      61  CGGAATCGCGTTGGCGGCAACATGTTGGCCGCCCGACGTTCCGTCGCCCGAGCAGCGC 120
Db      1039  CGACAGCACCGCGGAGGCCAGCTTGGGTACTGGCGGTGCGCCAGTTCGGCTCCTTGCC 1098

Qy      121  CGGCCCCAACGGCAGCGCGACCTCG 145
Db      1099  GGGCGCCGACCAGCAGCAGCGGCTCG 1123

```

Search completed: February 10, 2005, 06:44:45
Job time : 96 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 03:46:41 ; Search time 2098 Seconds
(without alignments)
5575.383 Million cell updates/sec

Title: US-10-647-089-93
Perfect score: 321
Sequence: 1 ttgacccacagcgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 32822875 seqs, 18219865908 residues

Total number of hits satisfying chosen parameters: 65645750

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:*

1: gb_est1:*	2: gb_est2:*	3: gb_hic:*	4: gb_est3:*	5: gb_est4:*	6: gb_est5:*	7: gb_est6:*	8: gb_gsl:*	9: gb_gsl2:*
--------------	--------------	-------------	--------------	--------------	--------------	--------------	-------------	--------------

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	49.8	15.5	925	9	CNS0091P	AL053013 Drosophila
2	46.8	14.6	627	7	CF481544	CF481544 POLI 72 D
3	45.8	14.6	792	7	CF487459	CF487459 POLI 43 H
4	45.2	14.1	626	6	CD927090	CD927090 GR45.100N
5	45	14.0	925	9	CNS0091P	AL053013 Drosophila
6	44.8	14.0	1101	9	CNS017SY	AL108460 Drosophila
7	43.6	13.6	532	9	P233L	AL160581 Leishmani
8	43.6	13.6	623	9	P305L	AL160634 Leishmani
9	43	13.4	478	6	CB029421	CB029421 TgESTzvd6
10	42.8	13.3	1075	8	BZ574121	BZ574121 msh2_3520
11	42.8	13.3	1331	8	BZ576260	BZ576260 msh2_4849
12	42.8	13.3	1348	8	BZ574088	BZ574088 msh2_3505
13	42.6	13.3	932	9	CNS0072Q	AL066742 Drosophila
14	42.4	13.2	514	6	CA248787	CA248787 SCBFL110
15	42.4	13.2	809	9	CC626073	CC626073 OGUHF43TV
16	42.4	13.2	906	9	CG178817	CG178817 PUFQ35TB
17	42.4	13.2	912	9	CG178819	CG178819 PUFQ35TB
18	42.4	13.2	958	9	CG084181	CG084181 FUIDR06TB
19	42.4	13.2	960	9	CC667226	CC667226 OGUET33TV
20	42.2	13.1	734	7	CF482703	CF482703 POLI 8 D0
21	42	13.1	323	4	BI776880	BI776880 EBPi03_SQ
22	42	13.1	429	5	BU983832	BU983832 HA31C2xr
23	42	13.1	530	6	CD895551	CD895551 G174.100C
24	42	13.1	638	6	CD902568	CD902568 G356.107G

C 25	42	13.1	658	5	BQ805102	BQ805102 WHE3562 H
C 26	42	13.1	659	6	CD885865	CD885865 G118.100L
C 27	42	13.1	747	9	CNS02PY0	AL208593 Tetraodon
C 28	41.8	13.0	1673	5	BU907501	BU907501 AGENCOURT
C 29	41.6	13.0	611	6	CD938304	CD938304 OV.109KL7
C 30	41.6	13.0	741	7	CF486940	CF486940 POLI 40 F
C 31	41.6	13.0	846	9	CNS010RJ	AL099337 Drosophila
C 32	41.2	12.8	780	8	BH370998	BH370998 AG-ND-170
C 33	41.2	12.8	844	9	CNS0052P	AL056652 Drosophila
C 34	41.2	12.8	1101	9	CNS01720	AL107514 Drosophila
C 35	41.2	12.8	1237	7	CK211563	CK211563 FGAS02341
C 36	41	12.8	400	1	AA443087	AA443087 zx74a02.s
C 37	40.8	12.7	673	6	CA221306	CA221306 SCSGFL403
C 38	40.8	12.7	942	5	BU534531	BU534531 AGENCOURT
C 39	40.6	12.6	1254	9	AG057481	AG057481 Pan trogl
C 40	40.6	12.6	1542	9	AG032943	AG032943 Pan trogl
C 41	40.4	12.6	704	9	AG060135	AG060135 Pan trogl
C 42	40.4	12.6	802	9	AG132220	AG132220 Pan trogl
C 43	40.2	12.5	537	4	BI804974	BI804974 S003B03 S
C 44	40.2	12.5	839	9	CNS004NB	AL054280 Drosophila
C 45	40.2	12.5	951	6	CD492951	CD492951 CDA02-D06

ALIGNMENTS

RESULT 1
CNS0091P 925 bp DNA linear GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR19D16 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION AL053013 GI:4934461
VERSION AL053013.1
KEYWORDS GSS.
SOURCE Drosophila melanogaster (fruit fly)
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 925)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see <http://www.fruitfly.org> The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammos in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial, EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's p1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
source location/Qualifiers
1..925
/organism="Drosophila melanogaster"
/mol_type="genomic DNA"
/db_xref="taxon:7227"
/clone="BACR19D16"
/clone_lib="RPCI-98"
/note="end : TET3"
ORIGIN
Query Match 15.5% ; Score 49.8 ; DB 9 ; Length 925 ;
Best Local Similarity 14.7% ; Pred. No. 0.024 ;

Matches	45;	Conservative	138;	Mismatches	123;	Indels	0;	Gaps	0;
Qy	5	CCACAAGCGCAGCTAAACGCCGAGCCAGCCATCGCGCAGGAGGCTCAACGCCCCGGGTGGGA	64						
Db	574	CCSSCSGSSSCBCCCCCSCSSYCCSSSBSSSKSSSTSBSSCCSCCCKSVKSGTSCSS	633						
Qy	65	ATCGCGTTGGCGGCAACATGTTGGCCGCGCAGCGTTCCGTCGCGCAGCAGCGCCGCG	124						
Db	634	SSSCSSSSSTSSSTSSSTSSSKSSSGSSSSSSSYTTKSTASGSGSNAGGSGSGTG	693						
Qy	125	CCCAAAGCGCAGCGGACCTGAGGCTATCCGCGAGCGTACGCCGAGATGGTGGCGCAT	184						
Db	694	STSSSSSSSTSTSSSVSGSKSTBSSGSSGSSSSSTSSBBSCTSTSSSSSSSY	753						
Qy	185	CACACAGAAATCGACGACACAGCCAACTGGCGCTGTTGTCGATGCATCTCGACCATG	244						
Db	754	SSSTCCTCCCSYSSSSSTSSSSSTSWGSTSGSSSVGTSSSSSTSTSTCCSCCCTMCT	813						
Qy	245	AGCAGCGCCCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCGGTATCACTTCCCGCAGC	304						
Db	814	CCSTYBMBCTYSTCGGSSSSSGKGVTKCGCGCGSSSTNGMBGTSACSSSSSCSS	873						
Qy	305	AAACCG	310						
Db	874	SVSSSS	879						

RESULT 2	CF481544/c	627 bp	mrna	linear	EST 08-SEP-2003
LOCUS	POLL172.D01.g1.A002	Pollen Sorghum	bicolor	cdna clone	
DEFINITION	POLL172.D01.A002 5', mRNA sequence.				
ACCESSION	CF481544				
VERSION	CF481544.1	GI:34510413			
KEYWORDS	EST				
SOURCE	Sorghum bicolor (sorghum)				
ORGANISM	Sorghum bicolor				
	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.				
REFERENCE	1 (bases 1 to 627)				
AUTHORS	Cordonnier-Pratt,M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J., Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and Pratt,L.H.				
TITLE	EST database from Sorghum: pollen				
JOURNAL	Unpublished (2003)				
COMMENT	Other ESTs: POLL172.D01.b1.A002				

Library for Genomics and Bioinformatics
The University of Georgia, Department of Plant Biology
Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA
Tel: 706 542 1860
Fax: 706 583 0210
Email: mmpratt@uga.edu

Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in
the Human Genome Center, University of Tokyo. Institute of Medical
Science; plant material and RNA prepared at Texas A & M University,
sequencing done in the Laboratory for Genomics and Bioinformatics,
University of Georgia. Sequence ends have been trimmed to exclude
vector and regions below Phred quality 16. Three-prime sequences
are presented as their reverse complement and have been trimmed to
exclude polyA.

```

Accession: F0141
Seq primer: Sug5 (CTTCTGCTCTTAAAGTCGC).
Location/Qualifiers
    1..627
    /organism="Sorghum bicolor"
    /mol_type="mRNA"
    /cultivar="BTx623"
    /db_xref="taxon:4558"
    /clone="POL1 72 D01_A002"
    /lab_host="DH10B-T1 phage-resistant E. coli"
    /clone_lib="Pollen"
FEATURES
    source

```

/note="Organ: Pollen; Vector: pME18S-FL3; Site 1: XhoI; Site 2: XhoI; The library was prepared from polyA⁺ RNA from pollen at the late vacuolated-vacuolated stage of development. Pollen was harvested from greenhouse-grown panicles of sorghum line Brx623. Panicles were removed from the flag leaf prior to emergence, when no detectable amylase is present in pollen of male-fertile lines. This stage represents pollen collected from anthers about 8-14 days prior to anthesis. Double-stranded cDNA was cloned unidirectionally into different DraIII sites of the pME18S-FL3 vector (5-prime DraIII site is CACTGTGG, 3-prime DraIII site is CACCATGTG). XhoI excises the cDNA insert."

ORIGIN

	Query Match	14.6%;	Score 46.8;	DB 7;	Length 627;
	Best Local Similarity	50.4%;	Pred. No. 0.14;		
	Matches 114;	Conservative 0;	Mismatches 112;	Indels 0;	Gaps 0;
QY	36	CGCCGCGAGGCTCAACGCGCCGCTCGGAATCGTTGGCGGCACATGTTGGCGCGC	95		
Db	519	CGCCGCGCTGCTGCTGCTGCGGTGGCATCGGTGGCCGCGTGTGAGCTTGGCGCGC	450		
QY	96	CGAGGTTTCGTCGCGGAGCAGCGCGCGCCCAACGGCAGCGGACCTCTGAGGCTATCCG	155		
Db	459	AGAGCGCCAGGTGGTGGAGCAGCGCGGAGAGCTCGGCCGTCCTCGATCTCGAGTGGCGCT	400		
QY	156	CCGAGGGTACGCGGAGATGGTCGGCGACATCACACGAAATCGACGACACAGCCGAACCT	215		
Db	399	CLAGGCTGAGGATGAGCGCGGTAGACGCCCGACGCGGGTCTTGGCGCGCGATGTCGGACT	340		
QY	216	GGCGCTGTGTTCATGTCATCTCGACGATGACGAGCGCCGGCTTGAG	261		
Db	339	GGTAGATGAGCGTGGCCATGGCGGTGGCGCAGAGCTCCGGGCTTGAG	294		

RESULT 3
CF487459/C

LOCUS	CF487459	792 bp	mRNA	linear	EST 08-SEP-2003
DEFINITION	POLI_43 H12.g1_A002 Pollen sorghum bicolor cDNA clone				
ACCESSION	POLI_43 H12_A002 5', mRNA sequence.				
VERSION	CF487459				
KEYWORDS	CF487459.1 GI:34516328				
SOURCE	EST.				
ORGANISM	Sorghum bicolor (sorghum)				
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD clade; Panicoideae; Andropogoneae; Sorghum.				
AUTHORS	1 (bases 1 to 792) Cordonnier-Pratt M.-M., Suzuki,Y., Sugano,S., Klein,R.R., Liang,C., Sun,F., Sullivan,R., Eastman,A., Cannon,R., Kern,B., Morgan,J., Lucas,A., Al-Sheikh,A., Jones,V., Adibi,N., Owen,A., Gao,J. and Pratt,L.H.				
TITLE	EST database from Sorghum: pollen				
JOURNAL	Unpublished (2003)				
COMMENT	Other ESTs: POLI_43 H12.b1_A002 Contact: Cordonnier-Pratt MM Laboratory for Genomics and Bioinformatics The University of Georgia, Department of Plant Biology Plant Sciences Building, Rm. 2502, Athens, GA 30602-7271, USA Tel: 706 542 1860 Fax: 706 583 0210 Email: mmpratt@uga.edu Library constructed by Dr. Yutaka Suzuki and Dr. Sumio Sugano in the Human Genome Center, University of Tokyo Institute of Medical Science: plant material and RNA prepared at Texas A & M University; sequencing done in the Laboratory for Genomics and Bioinformatics, University of Georgia. Sequence ends have been trimmed to exclude vector and regions below Phred quality 16. Three-prime sequences are presented as their reverse complement and have been trimmed to exclude polyA. seq primer: Sug5 (CTTCTGCTCTAAAGCTGCG).				

4

[illegible]

```

mesh2_3520.x1 msh Pseudomonas aeruginosa genomic clone mesh2_3520,
Genomic survey sequence.
ACCESSION      BZ574121
VERSION        BZ574121.1  GI:27209182
KEYWORDS       GSS.
SOURCE         Pseudomonas aeruginosa
ORGANISM       Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1075)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library
J. Bacteriol. (2002) In press
Contact: Chris K. Raymond
Genome Center
University of Washington
Box 352145, Seattle, WA 98105-2145, USA
Tel: 2062216954
Fax: 2066857244
Email: craymond@u.washington.edu
Class: shotgun.
Location/Qualifiers
1..1075
/organism="Pseudomonas aeruginosa"
/mol_type="genomic DNA"
/strain="WSH"
/db_xref="taxon:287"
/clone="mesh2_3520"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

ORIGIN
Query Match      13.3%; Score 42.8; DB 8; Length 1075;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY      48  CAACGCCCGCGTGGGAATCGGTTGGGGCGCAACATGTTGCCGCGCGAGTTCCGTC 107
Db      392  CCACGTCCTGCGACGAGAACGATGATCGGCTTCGCCGTTGGCGGCGACAGATGTTGTTCG 333
QY      108  CGCCGACGAGCGCCGCGCCCAACGGCAGCGACCTCGAGGGCTATTCGCCGAGCGGTACGC 167
Db      332  TGGACATCATCAGCGCGCGCGCTTCACAGCTGGCGCTCGAGGGTTCAGCGAACTGGACGG 273
QY      168  CGAGATGTTGGGCACATCACAGGAATCGACGACGACACAGCGGAAGTGGCGCTGTTGTC 227
Db      272  CCATCTGTGTCACCTCGGAAGTGGCGGTTGTACCGCGCGCAGACCAAGCGGTCGACGTGA 213
QY      228  GATGCATCTCGACGATGAGCAGCGCC 253
Db      212  TCGCCTTGCCTTCGATGAGGACCGC 187

RESULT 11
BZ576260/c
LOCUS
DEFINITION      mesh2_4849.y2 msh Pseudomonas aeruginosa genomic clone mesh2_4849,
Genomic survey sequence.
ACCESSION      BZ576260
VERSION        BZ576260.1  GI:27211321
KEYWORDS       GSS.
SOURCE         Pseudomonas aeruginosa
ORGANISM       Pseudomonas aeruginosa
Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
Pseudomonadaceae; Pseudomonas.
1 (bases 1 to 1331)
Spencer,D.H., Raymond,C.K., Smith,E.E., Sims,E.E., Hastings,M.,
Burns,J.L., Kaul,R. and Olsen,M.V.
Whole-Genome-Sequence variation among multiple isolates of
Pseudomonas aeruginosa library

```

```

/db_xref="taxon:287"
/clone="msh2.3505"
/clone_lib="msh"
/note="Environmental isolate. Whole genomic shotgun
library."

ORIGIN
Query Match      13.3%; Score 42.8; DB 8; Length 1348;
Best Local Similarity 50.5%; Pred. No. 1.4;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 48 CAACGCCCGGTCGGATCCGTTGGCGGCAACATGTTGGCGCGACGTTCCGTC 107
    |||||
Db 410 CCACGTCTCGACGGAACGATGATCGGCTCCGCTGGCGGCGACAGGATGTTGTGG 351
    |||||

QY 108 CGCGAGCAGCGCGCGCCCAACGCGCAGCGGACCTCGAGGCTATCCGCGGCGTACGC 167
    |||||
Db 350 TGGACATCATCAGCGCGCGCGTTCACGTTGGCTCGAGGTCAGCGGAACGTGGACG 291
    |||||

QY 168 CGAGATGTGGCGACATCACACGAAATCGACGACACAGCCGAACTGGCGCTTTGTC 227
    |||||
Db 290 CCATCTGTGTCACGTCGAAGTCGGCTTGTACGCGCGCGACAGCGGTCGACCTGA 231
    |||||

QY 228 GATGCATCTCGACGATGACGCGCC 253
    |||||
Db 230 TCGCCTTGCTTCGATGAGGACCGGC 205
    |||||

RESULT 13
CNS0072Q      932 bp DNA linear GSS 03-JUN-1999
LOCUS      Drosophila melanogaster genome survey sequence T7 end of BAC #
DEFINITION      BAC114B09 of RPC1-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION      AL066742
VERSION      AL066742.1 GI:4945205
KEYWORDS      GSS.
SOURCE      Drosophila melanogaster (fruit fly)
ORGANISM      Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
Ephydroidea; Drosophilidae; Drosophila.
REFERENCE      1 (bases 1 to 932)
AUTHORS      Direct Submission
TITLE      Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
JOURNAL      BP 191 91006 EVRY cedex - FRANCE (E-mail : segrengenoscope.cns.fr
COMMENT      - Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mammosser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPC1-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES      Location/Qualifiers
source      1..932
            /organism="Drosophila melanogaster"
            /mol_type="genomic DNA"
            /db_xref="taxon:7227"
            /clone="BAC114B09"
            /clone_lib="RPC1-98"
            /note="end : T7"

ORIGIN
Query Match      13.3%; Score 42.6; DB 9; Length 932;
Best Local Similarity 37.2%; Pred. No. 1.6;
Matches 51; Conservative 36; Mismatches 50; Indels 0; Gaps 0;

QY 22 GCCAGCCAGCATCGCGCAGGCTCAACGCCCGCTCGGAATCGGTTGGCGGCA 81
    |||||
Db 735 CSCCCSCCCSCGSSSGCGSGCGSGCGSGCGSGCGSGCGSGCGSGCGSG 794
    |||||

QY 82 CATGTTGGCGCGGAGTTCCTCCGTCGCGAGCAGCGCCCAACGCGCAGCGCGAC 141
    |||||
Db 795 CGSGSGCGCGSSSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSGSG 854
    |||||

QY 142 CTCGAGGCTATCCGCGC 158
    |||||
Db 855 SSCGCGSGCGSGSGCG 871
    |||||

RESULT 14
CA248787      514 bp mRNA linear EST 25-SEP-2003
LOCUS      SCSBFL1104G10.9 FL1 Saccharum officinarum cDNA clone SCSBFL1104G10
DEFINITION      5' mRNA sequence.
ACCESSION      CA248787
VERSION      CA248787.1 GI:35330788
KEYWORDS      EST.
SOURCE      Saccharum officinarum
ORGANISM      Saccharum officinarum
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD
clade; Panicoideae; Andropogoneae; Saccharum; Saccharum officinarum
complex.
REFERENCE      1 (bases 1 to 514)
AUTHORS      Vettore A.L., da Silva P.R., Kemper, E.L. and Arruda, P.
TITLE      The libraries that made SUCEST
JOURNAL      Genet. Mol. Biol. 24 (1-4), 1-7 (2001)
COMMENT      Contact: Arruda P
            Centro de Biologia Molecular e Engenharia Genetica
            Universidade Estadual de Campinas
            Caixa Postal 6010, 13083-970, Campinas SP, Brazil
            Tel: 55 19 3788 1137
            Fax: 55 19 3788 1089
            Email: parruda@unicamp.br
            Clone distribution: clone distribution information can be found
            through the Brazilian Clone Collection Center (BCCC) at
            http://www.bcccenter.fcav.unesp.br
            Plate: 104 row: G column: 10
            Seq primer: T7 Promoter Primer.
FEATURES      Location/Qualifiers
source      1..514
            /organism="Saccharum officinarum"
            /mol_type="mRNA"
            /db_xref="taxon:4547"
            /clone="SCSBFL1104G10"
            /lab_host="DH10B"
            /clone_lib="FL1"
            /note="Organ: Inflorescence at beginning of development
            (1cm-long); Vector: pSport1; Site_1: SalI; Site_2: NotI;
            An unidirectional cDNA library generated from
            (inflorescence at beginning of development (1cm-long)).
            cDNA was prepared from polyA+ mRNA using SuperScript
            Plasmid System Kit (Invitrogen). The double-strand cDNAs
            were fractionated in a sepharose CL-2B 40cm-columns and
            fragments sizing between 0.8 and 1.5 Kb were
            directionally cloned into the vector. Details
            of each
            source of RNA and library construction can be obtained at
            http://sucest.lad.ic.unicamp.br/public"

ORIGIN
Query Match      13.2%; Score 42.4; DB 6; Length 514;
Best Local Similarity 49.3%; Pred. No. 1.9;
Matches 109; Conservative 0; Mismatches 112; Indels 0; Gaps 0;

QY 92 CGGCGACGTTTCGTCGCGAGCAGCGCGCGCCCAACGCGCGCGACCTCGAGGCTA 151
    |||||

```

Search completed: February 10, 2005, 06:43:13
Job time : 2108 secs

```
Db      293  CGCCCGGAGCGCGCGGAGATCAACCTCTCCTGCGCGGAGCGCTTCGAGCGCG 352
QY      152  TCCGCGGAGCGTACCGGAGATGGTGGGACATCATCAGCAATCGACGACGACGCG 211
Db      353  TGGGCGCGCCCAAGCGCGGCGCTGGAGGCGCGCTGCCCGCGGTGGTCTCGCGCGAG 412
QY      212  AACTGGCGCTGTTGTGATGCATCTCGACGATGAGCAGCGCGGCTTGAGGGGGGATGA 271
Db      413  TCCTGGCCCTCGCGCGCGGACCTCTGGCCATCTGGCGGGCCACGGGTCCCCGTG 472
QY      272  AGCTCGGCTGGCATCGTATCATCTTCCCGAGAAACCGAC 312
Db      473  CGCTGGCGCGCGNGACTCCGCCGCTCCGACGCGCGGAC 513
```

RESULT 15

CC626073/c

LOCUS

DEFINITION OGUHF43TV ZM 0.7 1.5 KB. Zea mays genomic clone ZMMBma0439H14,
genomic survey sequence.

ACCESSION

CC626073

VERSION 1

KEYWORDS

SOURCE

ORGANISM

Zea mays

Zea mays

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD

clade; Panicoideae; Andropogoneae; Zea.

1 (bases 1 to 809)

Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,

Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,

Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

Consortium for Maize Genomics

Unpublished (2002)

Other_GSSs: OGUHF43TH

Contact: Cathy Whitelaw

TIGR

9712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TF

Classes: sheared ends.

FEATURES

source

1..809

/organism="Zea mays"

/mol_type="genomic DNA"

/strain="B73"

/db_xref="taxon:4577"

/clone="ZMMBma0439H14"

/clone_lib="ZM 0.7 1.5 KB"

/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb

methylation filtered genomic DNA library"

ORIGIN

Query Match

Best Local Similarity

Matches

88; Conservative

0; Mismatches

76; Indels

0; Gaps

0;

QY

96

CGACGTTCCGTCGCGGAGCGCGCGCCCAACGCGGAGCGGACCTCGAGGCTATCCG 155

Db

309

CGCGGATTGGGGGGGGCGCTGGGTTCTCGGACTCTCGCGGACGCGCGAGCCCGACGA 250

QY

156

CCGAGCGTACCGCGAGATGGTGGGACATCACAGAAATCGAGACGACAGCCGAACT 215

Db

249

GTTGGAGTGGCGCTGATGGTGGCGAGCTGCACAGAGCTCAAGGGGACCTCGCCGAGGC 190

QY

216

GGCGCTGTTGTCGATGCATCTCGACGATGACGAGCGCGCGCTTG 259

Db

189

TGAGCGTCACTTCAGGAGCTCTCGGAGAGGAGCCCTCTCTCG 146

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 03:44:31 ; Search time 335 Seconds
(without alignments)
5030.042 Million cell updates/sec

Title: US-10-647-089-93
Perfect score: 321
Sequence: 1 ttgaccacagcgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 413486 seqs, 2624710521 residues

Total number of hits satisfying chosen parameters: 8269772

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_23Sep04:*

- 1: Geneseqn1980s:*
- 2: Geneseqn1990s:*
- 3: Geneseqn2000s:*
- 4: Geneseqn2001as:*
- 5: Geneseqn2001bs:*
- 6: Geneseqn2002as:*
- 7: Geneseqn2002bs:*
- 8: Geneseqn2003as:*
- 9: Geneseqn2003bs:*
- 10: Geneseqn2003cs:*
- 11: Geneseqn2003ds:*
- 12: Geneseqn2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	321	100.0	324	4	Aaf24416 M tubercu
2	321	100.0	110000	4	Continuation (30 o
3	321	100.0	110000	4	Continuation (30 o
4	42.8	13.3	4200	4	Aaf51542 Pseudomon
5	42.8	13.3	4200	8	ACA19522 Prokaryot
6	42.8	13.3	4212	11	Abd06144 Pseudomon
7	42.8	13.3	4242	11	Abd08093 Pseudomon
8	42.8	13.3	10232	8	ACA37617 Prokaryot
9	41.6	13.0	996	8	ACA23398 Prokaryot
10	41	12.8	400	6	ABL81990 Human ova
11	40.4	12.6	951	8	ACA38286 Prokaryot
12	40.4	12.6	954	4	Aah51985 Mycobacte
13	40.4	12.6	954	8	ACA40430 Prokaryot
14	40.4	12.6	45190	6	Abx09144 Mycobacte
15	40.4	12.6	110000	4	Continuation (12 o
16	40.4	12.6	110000	4	Continuation (12 o
17	40.2	12.5	732	11	Abd07694 Pseudomon
18	40.2	12.5	945	11	Abd07563 Pseudomon
19	40.2	12.5	1602	11	Abd07778 Pseudomon
20	39.8	12.4	594	11	Abd13439 Pseudomon
21	39.8	12.4	966	11	Abd13292 Pseudomon

22	39.8	12.4	1179	11	ABD13413
23	39.6	12.3	1512	8	ABX71825
24	39.6	12.3	4509	8	ABX71813
25	39.6	12.3	4512	8	ABX71823
26	39.6	12.3	4767	3	AAA71571
27	39.4	12.3	1470	11	ABD07721
28	39.4	12.3	1473	11	ABD07540
29	39.4	12.3	1488	4	AAS54225
30	39.4	12.3	1488	8	ACA42505
31	39.4	12.3	1512	11	ABD07615
32	39.2	12.2	1320	8	ABX56068
33	38.8	12.1	408	11	ABD02051
34	38.8	12.1	603	8	ACA26989
35	38.8	12.1	801	11	ABD02094
36	38.8	12.1	1356	11	ABD02026
37	38.8	12.1	1434	8	ACA35463
38	38.8	12.1	2955	11	ABD02080
39	38.8	12.1	3150	11	ABD02123
40	38.6	12.0	109519	5	AAS08693
41	38.6	12.0	2379	4	AAS54231
42	38.6	12.0	2379	8	ACA42520
43	38.6	12.0	2379	8	ABT14638
44	38.4	12.0	1371	11	ABD15350
45	38.4	12.0	2181	11	ABD15138

ALIGNMENTS

RESULT 1
AAF24416
ID AAF24416 standard; DNA; 324 BP.
XX
AC AAF24416;
XX
DT 24-APR-2001 (first entry)
XX
DE M tuberculosis Rv2653c gene.
XX
KW Tuberculosis; TB; vaccine; esat-6 gene family; Rv0287; Rv1036c; Rv1037c;
KW Rv2346c; Rv2348c; Rv2653c; Rv2654c; Rv3020c; Rv3444c; Rv3445c; Rv3990c;
KW Rv3891c; Rv3904c; Rv3905c; ds.
XX
OS Mycobacterium tuberculosis.
XX
PN WO200104151-A2.
XX
PD 18-JAN-2001.
XX
PF 13-JUL-2000; 2000WO-DK000398.
XX
PR 13-JUL-1999; 99DK-00001020.
XX
PA 15-JUL-1999; 99US-0144011P.
XX
(STAT-) STATENS SERUM INST.
XX
PI Andersen P, Skjot R;
XX
DR WPI; 2001-091923/10.
XX
P-PSDB; AAB35226.
XX
PT Immunizing against and diagnosis of tuberculosis.
XX
PS Claim 28; Page 71; 80pp; English.
XX
CC The present invention provides the protein and coding sequences for
CC members of the esat-6 gene family from Mycobacterium tuberculosis. These
CC proteins include Rv0287, Rv1037c, Rv2346c, Rv2348c, Rv2653c,
CC Rv2654c, Rv3020c, Rv3444c, Rv3445c, Rv3891c, Rv3904c and
CC Rv3905c. These can be used to produce vaccines against, and in the
CC diagnosis of, tuberculosis (TB) infection. The present sequence is one of
CC the coding sequences of the invention

WP AAI99683_29 2900001 3010000
 WP AAI99683_30 3000001 3110000
 WP AAI99683_31 3100001 3210000
 WP AAI99683_32 3200001 3310000
 WP AAI99683_33 3300001 3410000
 WP AAI99683_34 3400001 3510000
 WP AAI99683_35 3500001 3610000
 WP AAI99683_36 3600001 3710000
 WP AAI99683_37 3700001 3810000
 WP AAI99683_38 3800001 3910000
 WP AAI99683_39 3900001 4010000
 WP AAI99683_40 4000001 4110000
 WP AAI99683_41 4100001 4210000
 WP AAI99683_42 4200001 4310000
 WP AAI99683_43 4300001 4403765

Query Match 100.0%; Score 321; DB 4; Length 110000;
 Best Local Similarity 100.0%; Pred. No. 9.9e-70;
 Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TTGACCCACAGCGCACTAAAGCCAGCCAGCCATCGCGCGAGGCTCAACGCCCGCGGT 60
 DB 71654 TTGACCCACAGCGCACTAAAGCCAGCCAGCCATCGCGCGAGGCTCAACGCCCGCGGT 71595

QY 61 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGGACGTTCCGTCGCGGAGCAGCGC 120
 DB 71594 CGGAATCGGTTGGCGCGCAACATGTTGGCGCGGACGTTCCGTCGCGGAGCAGCGC 71535

QY 121 CGCGCCCAAGCGCAGCGCACTCGAGGCTATCGCGCGAGCGTAGCGCGAATGTCGCG 180
 DB 71534 CGCGCCCAAGCGCAGCGCACTCGAGGCTATCGCGCGAGCGTAGCGCGAATGTCGCG 71475

QY 181 ACATCACAGCAATCGACGACGACACGACCGCACTGGCGCTGTTGTCGATCATCTCGAC 240
 DB 71474 ACATCACAGCAATCGACGACGACACGACCGCACTGGCGCTGTTGTCGATCATCTCGAC 71415

QY 241 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTTATCATTTCC 300
 DB 71414 GATGAGCAGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTTATCATTTCC 71355

QY 301 GACGAACCCGACGACCAACAG 321
 DB 71354 GACGAACCCGACGACCAACAG 71334

RESULT 4
 AAS51542/c
 ID AAS51542 standard; DNA; 4200 BP.
 AC AAS51542;
 XX
 XX
 DT 13-FEB-2002 (first entry)
 XX
 DE Pseudomonas aeruginosa DNA for cellular proliferation protein #127.
 KW Antisense; ds; prokaryotic cellular proliferation gene; antibiotic;
 KW antibacterial; drug design.
 XX Pseudomonas aeruginosa.
 OS
 XX WO200170955-A2.
 XX
 XX
 PD 27-SEP-2001.
 XX
 XX 21-MAR-2001; 2001WO-US009180.
 XX
 XX 21-MAR-2000; 2000US-0191078P.
 PR 23-MAY-2000; 2000US-0206848P.
 PR 26-MAY-2000; 2000US-0207272P.
 PR 27-OCT-2000; 2000US-0242578P.
 PR 27-NOV-2000; 2000US-0253625P.
 PR 22-DEC-2000; 2000US-0257931P.
 PR 16-FEB-2001; 2001US-0269308P.

(ELIT-) ELITRA PHARM INC.
 Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
 Yamamoto RT, Xu HH;
 WPI; 2001-611495/70.
 P-PSDB; AAU33683.

New polynucleotides for the identification and development of
 antibiotics, comprise sequences of antisense nucleic acids.

Claim 27; SEQ ID NO 4124; 511pp; English.

The invention relates to antisense inhibitors of genes essential to
 prokaryotic cellular proliferation, their use in identifying the genes,
 their use in the discovery of novel antibiotics, the essential genes,
 themselves and the encoded proteins. The prokaryotes used are Escherichia
 coli, Staphylococcus aureus, Salmonella typhi, Klebsiella pneumoniae,
 Pseudomonas aeruginosa and Enterococcus faecalis. The invention is also
 useful for the identification of potential new targets for antibiotic
 development. The antisense nucleic acids can also be used to identify
 proteins used in proliferation, to express these proteins, and to obtain
 antibodies capable of binding to the expressed proteins. The proteins can
 be used to screen compounds in rational drug discovery programmes. The
 antisense nucleic acid sequence is also useful to screen for homologous
 nucleic acids which are required for cell proliferation in a wide variety
 of organisms. The present sequence encodes an essential prokaryotic
 cellular proliferation protein. Note: The sequence data for this patent
 did not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

XX
 SQ Sequence 4200 BP; 846 A; 1329 C; 1304 G; 721 T; 0 U; 0 Other;
 Query Match 13.3%; Score 42.8; DB 4; Length 4200;
 Best Local Similarity 50.5%; Pred. No. 0.42;
 Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;

QY 48 CAACGCCCGCGTGGGAATCGGTTGGCGCGCAACATGTTGGCGCGGAGTTCCGTC 107
 DB 1519 CCAGTCTCGACGGAACGATGATCGGCTCGCGGCGGACAGGATGTTGTTGG 1460

QY 108 CGCGAGCAGCGCGCGCCCAACGCGCAGCGGACCTCGAGGCTATCCGCGGAGCGTACGC 167
 DB 1459 TGGACATCATCAGCGCGCGCGCTTCCAGCTGGGCTCGAGGCTCAGCGGAACGTGACGG 1400

QY 168 CGAGATGTCGCGACATCACACGAAATCGACGACGACACAGCGGAACCTGGCGCTGTTGC 227
 DB 1399 CCATCTGTCACCGTCGAAGTCGGCGTTGTACGCGCGCAGACACCGCGGTGTCAGCTGGA 1340

QY 228 GATGCATCTCGACGATGACGCGCC 253
 DB 1339 TCGCCTTCCTCGATGAGGACCGGC 1314

RESULT 5
 ACA19522/c
 ID ACA19522 standard; DNA; 4200 BP.
 AC ACA19522;
 XX
 XX 19-JUN-2003 (first entry)
 XX
 DE Prokaryotic essential gene #1179.
 KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 KW drug design; gene.
 XX Pseudomonas aeruginosa.
 OS
 XX WO200277183-A2.
 PN
 XX

PD 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
PF 21-MAR-2001; 2001US-00815242.
XX 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362659P.
XX (ELIT-) ELITRA PHARM INC.
PA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
XX Wali D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
PI WPI; 2003-029926/02.
XX P-PSDB; ABU15652.
DR New antisense nucleic acids, useful for identifying proteins or screening
XX for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
PT
XX Claim 14; SEQ ID NO 7392; 1766pp; English.
XX
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
CC prokaryotic essential genes. Note: The sequence data for this patent did
CC not form part of the printed specification, but was obtained in
CC electronic format directly from WIPO at
XX ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 4200 BP; 846 A; 1329 C; 1304 G; 721 T; 0 U; 0 Other;
Query Match 13.3%; Score 42.8; DB 8; Length 4200;
Best Local Similarity 50.5%; Pred. No. 0.42;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 48 CAACGCCCGCGCTCGGAATCGCTGGCGGCAACATGTTGGCGCGCGAGTTCGTC 107
DB 1519 CCACGTCTCTGCGACGAGACGATGATCGCTCGCGCTGGCGGCGACAGATGTTGTTGG 1460
QY 108 CGCCGAGCAGCGCGCGCGCCCAACCGGAGCGCGCACTTCGAGGCTATCCGCGAGCGTACGC 167
DB 1459 TGGACATCATCAGCGCGCGCGCTTCAGCTGGGCTCGAGGCTCAGCGAAGCTGAGCGG 1400
QY 168 CGAGATGTTGGCGACATCATCAGAAATCAGACGACGACAGCGGAACTCGCGCTGTTGTC 227
DB 1399 CCATCTGGTACCGTCAAGATCGGCGCTTGTACGCGCGCAGACGACGCGGTCAGTGA 1340

QY 228 GATGCATCTCGACGATGACGAGCGCC 253
DB 1339 TCGCCTTGCTTCGATGAGGACCGGC 1314
RESULT 6
ABD06144/C
ID ABD06144 standard; DNA; 4212 BP.
XX AC ABD06144;
XX DT 29-JUL-2004 (first entry)
XX DE Pseudomonas aeruginosa polynucleotide #4748.
XX Bacterial infection; gene; ds; Pseudomonas aeruginosa infection;
KW antibacterial.
XX Pseudomonas aeruginosa.
XX OS
XX PN US6551795-B1.
XX 22-APR-2003.
XX PD 18-FEB-1999; 99US-00252991.
XX PF 18-FEB-1998; 98US-0074788P.
XX PR 27-JUL-1998; 98US-0094190P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Rubenfield MJ, Nolling J, Deloughery C, Bush D;
XX WPI; 2003-615309/58.
XX P-PSDB; ABO72573.
XX Novel isolated nucleic acid encoding Pseudomonas aeruginosa polypeptide,
PT useful as molecular targets for diagnostics, prophylaxis and treatment of
PT pathological conditions resulting from bacterial infection.
XX Disclosure; SEQ ID NO 4748; 455pp; English.
XX The invention relates to Pseudomonas aeruginosa polypeptides and the
CC polynucleotides encoding them. The sequences are useful in diagnosis and
CC therapy of pathological conditions, as molecular targets for diagnostics,
CC prophylaxis and treatment of pathological conditions resulting from a
CC bacterial infection, for evaluating a compound, such as a polypeptide,
CC for the ability to bind a *P. aeruginosa* nucleic acid, as components of
CC effective antibacterial targets, as targets for antibacterial drugs,
CC including anti-*P. aeruginosa* drugs, as templates for recombinant
CC production of *P. aeruginosa*-derived peptides or polypeptides, as target
CC components for diagnosis and/or treatment of *P. aeruginosa*-caused
CC infection, and in detection of *P. aeruginosa* sequences or other sequences
CC of Pseudomonas species using biochip technology. Sequences ABD01397-
CC ABD17967 represent *P. aeruginosa* polynucleotides of the invention. Note:
CC The sequence data for this patent did not form part of the printed
CC specification but was obtained in electronic format from USPTO at
XX seqdata.uspto.gov/sequence.html
XX
SQ Sequence 4212 BP; 850 A; 1338 C; 1311 G; 713 T; 0 U; 0 Other;
Query Match 13.3%; Score 42.8; DB 11; Length 4212;
Best Local Similarity 50.5%; Pred. No. 0.42;
Matches 104; Conservative 0; Mismatches 102; Indels 0; Gaps 0;
QY 48 CAACGCCCGCGCTCGGAATCGCTGGCGGCAACATGTTGGCGCGCGAGTTCGTC 107
DB 1531 CCACGTCTCTGCGACGAGACGATGATCGCTCGCGCTGGCGGCGACAGATGTTGTTGG 1472
QY 108 CGCCGAGCAGCGCGCGCGCCCAACCGGAGCGCGGCACTTCGAGGCTATCCGCGAGCGTACGC 167
DB 1471 TGGACATCATCAGCGCGCGCGCTTCAGTGGGCTTCAGGCTCAGCGAAGCTGAGCGG 1412

CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: the sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 10232 BP; 1443 A; 3537 C; 3647 G; 1605 T; 0 U; 0 Other;

Query Match 13.3%; Score 42.8; DB 8; Length 10232;
 Best Local Similarity 47.7%; Pred. No. 0.45;
 Matches 125; Conservative 0; Mismatches 137; Indels 0; Gaps 0;
 QY 20 AACGCGAGCCAGCATCGCCGAGGCTCAACGCCCGCTCGAATCGGTGGCGGC 79
 Db |||||
 1271 AACTGTGACCGGAGAGACCCGCGCTGATCAGCCGGTGAGCAACCGTCCGCGCTGA 1330
 QY 80 AACATGTTGGCGGGCGAGCTTCGCTCCGCGAGCAGCGCCGCCCAACGCGAGCGG 139
 Db |||||
 1331 CCGAGCCCGACCCGCGCGCTGCTGATCCCGAGGCTTCGCGAACAGTGCAGCGCA 1390
 QY 140 ACTCGAGGCTATCGCCGAGCGTACCGGAGATGGTGGCGCATCACAGCAATGACG 199
 Db |||||
 1391 CCCCGCAGCGGTGGCGGTGACGTTCCGGGCGACCTCGCTGACCTACGCGCCAGCTCGAG 1450
 QY 200 ACACACAGCCGAACTGGCGCTGTGTCGATGATCTCGACGATGAGCGCGCGCTTG 259
 Db |||||
 1451 AGCGTCCACCGCTGGGCCATCTGCTCCCGACACGCGTGGCGCGGCGACTGCG 1510
 QY 260 AGCGGGGATGAAGCTCGGCTG 281
 Db |||||
 1511 TGGCGGTGATGTTCCCGCGTG 1532

RESULT 9

AC23398

ID ACA23398 standard; DNA; 996 BP.

XX ACA23398;

AC ACA23398;

XX ACA23398;

DT 19-JUN-2003 (first entry)

XX Prokaryotic essential gene #5055.

DE Antisense; ds; prokaryotic essential gene; cell proliferation;

XX Antisense; ds; prokaryotic essential gene; cell proliferation;

KW drug design; gene.

XX Borrelia cepacia.

OS WO200277183-A2.

XX WO200277183-A2.

PN 03-OCT-2002.

XX 03-OCT-2002.

FF 21-MAR-2002; 2002WO-US009107.

XX 21-MAR-2002; 2002WO-US009107.

PR 21-MAR-2001; 2001US-00815242.

XX 21-MAR-2001; 2001US-00815242.

PR 06-SEP-2001; 2001US-00948993.

XX 06-SEP-2001; 2001US-00948993.

PR 25-OCT-2001; 2001US-0342923P.

XX 25-OCT-2001; 2001US-0342923P.

PR 08-FEB-2002; 2002US-00072851.

XX 08-FEB-2002; 2002US-00072851.

PR 06-MAR-2002; 2002US-03626599.

XX 06-MAR-2002; 2002US-03626599.

PA (ELIT-) ELITRA PHARM INC.

XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;

XX Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;

DR WPI: 2003-029926/02.
 DR P-PSDB; ABU19528.
 XX New antisense nucleic acids, useful for identifying proteins or screening
 PT for homologous nucleic acids required for cellular proliferation to
 PT isolate candidate molecules for rational drug discovery programs.
 XX Claim 14; SEQ ID NO 11268; 1766pp; English.

XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,
 CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences

SQ Sequence 996 BP; 138 A; 320 C; 387 G; 151 T; 0 U; 0 Other;

Query Match 13.0%; Score 41.6; DB 8; Length 996;

Best Local Similarity 50.8%; Pred. No. 0.73;

Matches 127; Conservative 0; Mismatches 119; Indels 4; Gaps 1;

QY 29 CAGCCATCGCCGAGGCTCAACGCCCGCTCGAATCGGTGGCGGCACATGTT 88

Db |||||

145 CTGCAACGCGCGCGCTCGCTCGATCGATCGCGCTCGCGCTCGCGAGCGTTGGGAT 204

QY 89 GCGCGCGCGAGCTTCGCTCCGCGAGCAGCGCGCGCCCA-----CGCAGCGCGACCTC 144

Db |||||

205 CGCCAGCGCGAGCTTCGTCAGCGCTCGATCGCGCGCGAGCGCTGGTGGCGACGCGC 264

QY 145 GAGGCTATCGCGCGAGCGTACCGCGAGATGTTGGCGCATCATCAGCAATCGACACGAC 204

Db |||||

265 GACGGCTGTCCGCTCCGCTCGCGCTCGAGCGCGAGTGGCGGACAGACATTCCCGACG 324

QY 205 ACAGCGCAACTGGCGCTGTTGTCGATGTCATCTCGACGATGACGACGCGCGCTTGAGGG 264

Db |||||

325 CGCGCGCATCTGGAAGCGGTGAAGCAGTCTGTCGACGATGTCGGCAACCGCTGAACGCG 384

QY 265 GGGATGAAGC 274

Db |||||

385 TTGTCGACG 394

RESULT 10

ABL81990

ID ABL81990 standard; cDNA; 400 BP.

XX ABL81990;

AC ABL81990;

XX DT 17-MAY-2002 (first entry)
 XX AC Human ovarian cancer related cDNA clone SEQ ID NO:4968.
 XX DE Human; ovarian cancer; ovarian tumour; cytostatic; gene; ss.
 XX KW Homo sapiens.
 XX OS WO200192581-A2.
 XX PN 06-DEC-2001.
 XX PD 29-MAY-2001; 2001WO-US017756.
 XX PF 26-MAY-2000; 2000US-0207484P.
 XX PR (CORI-) CORIXA CORP.
 XX PA Algate PA, Harlocker SL, Jones R;
 XX PI WPI; 2002-122075/16.
 XX DR Composition for therapy and diagnosis of ovarian cancer comprising
 XX PT polypeptide of a ovarian tumor polypeptide, polynucleotide encoding
 XX PT polypeptide, antibody specific to polypeptide or T cell expressing
 XX PT polypeptide.
 XX PS Claim 1; SEQ ID NO 4968; 489pp; English.
 XX PS The present invention describes a composition (I) comprising: carriers
 CC and immunostimulants; and a polypeptide (II) of a ovarian tumour
 CC polypeptide encoded by a polynucleotide (III) having a cDNA sequence (S1)
 CC from the 10912 nucleotide sequences as given in ABL77023 to ABL87934.
 CC (III) encoding (II) having a sequence (S2), a T cell population of (II),
 CC or antigen presenting cells that express (II). (I) has cytostatic
 CC activity. An oligonucleotide (IV) that hybridises to (S1) can be used for
 CC detecting ovarian cancer in a patient's biological sample preferably
 CC serum or ovarian tissue. The method comprises contacting a biological
 CC sample from a patient with (IV), detecting the amount of polynucleotide
 CC hybridising to (IV) and comparing the amount to a predetermined cutoff
 CC value and thereby detecting ovarian cancer in the patient, where the
 CC amount of polynucleotide hybridising to (IV) is detected preferably by
 CC polymerase chain reaction (PCR). (I) comprising (III) and/or (II) is
 CC useful for stimulating and/or expanding T cells specific for an ovarian
 CC tumour protein comprising contacting T cells with (III) or (II). (III) is
 CC useful in design and preparation of ribozyme molecules for inhibiting
 CC expression of the tumour polypeptides and proteins in tumour cells; and
 CC to isolate a full length gene from a suitable library e.g., a tumour cDNA
 CC library using well known techniques
 XX SQ Sequence 400 BP; 57 A; 156 C; 135 G; 52 T; 0 U; 0 Other;
 Query Match 12.8%; Score 41; DB 6; Length 400;
 Best Local Similarity 58.7%; Pred. NO. 0.95;
 Matches 71; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
 Qy 90 GCCGCCGACGTTCCGTCGCGGACGACGCGCGGCCCAACGCGACGCGACCTCGAGGC 149
 Db 65 GCGGAAGAGGCGGCTCCACAGTGCAGACCTCGCGCGACGCGCGCGGCGCTCTCGC 124
 Qy 150 TATCCGCGGAGGTACGCGGAGATGGTGGCGACATCACAGAAATCGACGACGACGAC 209
 Db 125 CTCGCGGCTTGGCAGGCGGACACATCGCGCGGTATCGCTCGAAGAGGCGCGGAGTCTATC 184
 Qy 210 C 210
 Db 185 C 185
 RESULT 11
 ACA38286
 ID ACA38286 standard; DNA; 951 BP.

XX AC ACA38286;
 XX DT 19-JUN-2003 (first entry)
 XX DE Prokaryotic essential gene #19943.
 XX KW Antisense; ds; prokaryotic essential gene; cell proliferation;
 XX KW drug design; gene.
 XX OS Mycobacterium bovis.
 XX PN WO200277183-A2.
 XX PD 03-OCT-2002.
 XX PF 21-MAR-2002; 2002WO-US009107.
 XX PR 21-MAR-2001; 2001US-00815242.
 XX PR 06-SEP-2001; 2001US-00948993.
 XX PR 25-OCT-2001; 2001US-0342923P.
 XX PR 08-FEB-2002; 2002US-00072851.
 XX PR 06-MAR-2002; 2002US-0362699P.
 XX (ELIT-) ELITRA PHARM INC.
 XX FA Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
 XX PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
 XX PI WPI; 2003-029926/02.
 XX DR P-PSDB; ABU34416.
 XX DR New antisense nucleic acids, useful for identifying proteins or screening
 XX PT for homologous nucleic acids required for cellular proliferation to
 XX PT isolate candidate molecules for rational drug discovery programs.
 XX PS Claim 14; SEQ ID NO 26156; 1766pp; English.
 XX The invention relates to an isolated nucleic acid comprising any one of
 CC the 6213 antisense sequences given in the specification where expression
 CC of the nucleic acid inhibits proliferation of a cell. Also included are:
 CC (1) a vector comprising a promoter operably linked to the nucleic acid
 CC encoding a polypeptide whose expression is inhibited by the antisense
 CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
 CC polypeptide or its fragment whose expression is inhibited by the
 CC antisense nucleic acid; (4) an antibody capable of specifically binding
 CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular
 CC proliferation or the activity of a gene in an operon required for
 CC proliferation; (7) identifying a compound that influences the activity of
 CC the gene product or that has an activity against a biological pathway
 CC required for proliferation, or that inhibits cellular proliferation; (8)
 CC identifying a gene required for cellular proliferation or the biological
 CC pathway in which a proliferation-required gene or its gene product lies
 CC or a gene on which the test compound that inhibits proliferation of an
 CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
 CC compound's activity; (11) a culture comprising strains in which the gene
 CC product is overexpressed or underexpressed; (12) determining the extent
 CC to which each of the strains is present in a culture or collection of
 CC strains; or (13) identifying the target of a compound that inhibits the
 CC proliferation of an organism. The antisense nucleic acids are useful for
 CC identifying proteins or screening for homologous nucleic acids required
 CC for cellular proliferation to isolate candidate molecules for rational
 CC drug discovery programs, or for screening homologous nucleic acids
 CC required for proliferation in cells other than S. aureus, S. typhimurium,
 CC K. pneumoniae or P. aeruginosa. The present sequence is one of the target
 CC prokaryotic essential genes. Note: The sequence data for this patent did
 CC not form part of the printed specification, but was obtained in
 CC electronic format directly from WIPO at
 CC ftp.wipo.int/pub/published_pct_sequences
 XX SQ Sequence 951 BP; 141 A; 308 C; 329 G; 173 T; 0 U; 0 Other;
 Query Match 12.6%; Score 40.4; DB 8; Length 951;

Best Local Similarity 50.4%; Pred. No. 1.4;
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;
QY 72 TGGGCGGCAACATGGTTGGCGCGCGAGCTTCGTCGCGAGAGCGCGCGCCCAACG 131
Db 276 TCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCGAGGTTTACCGACTAACGGTGGT 335
QY 132 GCAGCGGCACTCGAGGCTATCCCGCGAGCGTA-----CGCGAGATGTTGGCGACATC 185
Db 336 CAATCGGACGCTCTGCGCTGCGCGGAGGATCTAGCGCGCGCGCGCGCGCGTGGT 395
QY 186 ACAGCAATCGACGACGACGACGCGGAACTGGCGCTTTTTCGATGATCTCGACGATGA 245
Db 396 TGCCAAATCTGCGGTACAACTAGCGGTACCGCGCTTGTTCATCTGCTTGTGAGTTTCCC 455
QY 246 GCAGCGCGGCTTCAGCGGCGGATGAAGCTCGCTGCGATCCGATATCACTTCCCGACGA 305
Db 456 GTCGATCGGTGTCGAGCGGTGATGGTGCAGCGCGAGGTCGCCGAACGGCTCGCGCCGA 515
QY 306 ACCCGACGCAAAACAG 321
Db 516 GCGGCGCAGCAAGAG 531

RESULT 12
AAH51985
ID AAH51985 standard; DNA; 954 BP.
XX AC AAH51985;
DT 04-SEP-2001 (first entry)
XX Mycobacterium tuberculosis potential drug target gene SEQ ID 39.
DE Drug target; growth; organism viability; characterisation; ds.
XX Mycobacterium tuberculosis.
OS W0200135317-A1.
XX 17-MAY-2001.
XX 13-NOV-2000; 2000WO-US031152.
XX 12-NOV-1999; 99US-0165086P.
PR 12-NOV-1999; 99US-0165124P.
PR 01-FEB-2000; 2000US-0179531P.
XX (REGC) UNIV CALIFORNIA.
XX Eisenberg D, Rotstein SH, Marcotte EM;
PI WPI; 2001-329193/34.
DR P-PSDB; AAG81134.
XX Identifying nucleotide or polypeptide sequence for use as drug target,
PT involves providing algorithm that analyzes a functional relationship
PT between nucleotide or polypeptide sequences, and comparing the sequences.
XX Disclosure; Page 76-77; 207pp; English.

This invention relates to a method for identifying a nucleotide or
CC polypeptide sequence that may be a drug target, or essential for growth
CC or viability of an organism. Polynucleotide sequences AAH51947 - AAH52092
CC represent DNA encoding proteins AAG81096 - AAG81241, Mycobacterium
CC tuberculosis proteins which are potential drug targets. The DNA and
CC protein sequences are used to illustrate the method of the invention. The
CC method involves providing an unknown nucleotide or polypeptide sequences,
CC and comparing it to a number of sequences along with at least one
CC algorithm capable of analysing a functional relationship between
CC nucleotide and polypeptide sequences. The method is useful for
CC characterising the function of nucleic acids and polypeptides that may be
CC useful as a target for a drug or essential for the growth or viability of

CC an organism
XX SQ Sequence 954 BP; 142 A; 309 C; 330 G; 173 T; 0 U; 0 Other;
Query Match 12.6%; Score 40.4; DB 4; Length 954;
Best Local Similarity 50.4%; Pred. No. 1.4;
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;
QY 72 TGGGCGGCAACATGGTTGGCGCGCGAGCTTCGTCGCGAGAGCGCGCGCCCAACG 131
Db 276 TCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCGAGGTTTACCGACTAACGGTGGT 335
QY 132 GCAGCGGCACTCGAGGCTATCCCGCGAGCGTA-----CGCGAGATGTTGGCGACATC 185
Db 336 CAATCGGACGCTCTGCGCTGCGCGGAGGATCTAGCGCGCGCGCGCGCGTGGT 395
QY 186 ACAGCAATCGACGACGACGACGCGGAACTGGCGCTTTTTCGATGATCTCGACGATGA 245
Db 396 TGCCAAATCTGCGGTACAACTAGCGGTACCGCGCTTGTTCATCTGCTTGTGAGTTTCCC 455
QY 246 GCAGCGCGGCTTCAGCGGCGGATGAAGCTCGCTGCGATCCGATATCACTTCCCGACGA 305
Db 456 GTCGATCGGTGTCGAGCGGTGATGGTGCAGCGCGAGGTCGCCGAACGGCTCGCGCCGA 515
QY 306 ACCCGACGCAAAACAG 321
Db 516 GCGGCGCAGCAAGAG 531

RESULT 13
ACA40430
ID ACA40430 standard; DNA; 954 BP.
XX AC ACA40430;
DT 19-JUN-2003 (first entry)
XX Prokaryotic essential gene #22087.
DE Antisense; ds; prokaryotic essential gene; cell proliferation;
KW drug design; gene.
XX Mycobacterium tuberculosis.
OS W0200277183-A2.
XX 03-OCT-2002.
XX 21-MAR-2002; 2002WO-US009107.
XX 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX (ELIT-) ELITRA PHARM INC.
XX Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX WPI; 2003-029926/02.
DR P-PSDB; ABU36560.

New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX Claim 14; SEQ ID NO 28300; 1766pp; English.
XX The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:

(1) a vector comprising a promoter operably linked to the nucleic acid encoding a polypeptide whose expression is inhibited by the antisense nucleic acid; (2) a host cell containing the vector; (3) an isolated polypeptide or its fragment whose expression is inhibited by the antisense nucleic acid; (4) an antibody capable of specifically binding the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular proliferation or the activity of a gene in an operon required for proliferation; (7) identifying a compound that influences the activity of the gene product or that has an activity against a biological pathway; (8) the gene product or that inhibits cellular proliferation; (9) identifying a gene required for cellular proliferation or the biological pathway in which a proliferation-required gene or its gene product lies or a gene on which the test compound that inhibits proliferation of an organism acts; (10) manufacturing an antibiotic; (11) profiling a compound's activity; (12) a culture comprising strains in which the gene product is overexpressed or underexpressed; (13) determining the extent to which each of the strains is present in a culture or collection of strains; or (14) identifying the target of a compound that inhibits the proliferation of an organism. The antisense nucleic acids are useful for identifying proteins or screening for homologous nucleic acids required for cellular proliferation to isolate candidate molecules for rational drug discovery programs, or for screening homologous nucleic acids required for proliferation in cells other than *S. aureus*, *S. typhimurium*, *K. pneumoniae* or *P. aeruginosa*. The present sequence is one of the target prokaryotic essential genes. Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences

Query Match 12.6%; Score 40.4; DB 8; Length 954;
Best Local Similarity 50.4%; Pred. No. 1.4;
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 72 TGGGGGCAACATGTTGGCGCGCGAGCTTCGGTCCGCGGAGAGCGCGCGCCCAACG 131
DB 276 TCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCGAGGTTACCGCACTAACCGTGGT 335
QY 132 GCAGCGCGACCTCGAGGCTATCCCGCGAGCGTA-----CGCCGAGATGGTGGGACATC 185
DB 336 CAATCGCGACGCTCTGGCCCTGCGCGGAGAGATCTAGCGCGGCGCGCGCGGTTGGT 395
QY 186 ACAGAAATCGACAGACGACGCGGAGTTCGGTTCGATCGATCTCGACGATGA 245
DB 396 TGCCAATCTGCGGTACACGATAGCGGTACCGCGGTTGTCATCTGTCGAGTTCC 455
QY 246 GCAGCGCGGTTGAGCGGAGTGAAGCTCGGCTGGGATCGGTATCACTTCCCGCGAGA 305
DB 456 GTCGATCGGTGTCGTCGCGGTGATGGTCAGCGCGAGGTGCGCGAACCGCTCGCGCGGA 515
QY 306 ACCGACAGCAACAG 321
DB 516 GCCGGGCAACAGAG 531

RESULT 14
ABX09144
ID ABX09144 standard; DNA; 45190 BP.

AC ABX09144;
XX
DT 08-APR-2003 (first entry)
XX
DE Mycobacterium tuberculosis H37Rv BAC clone BAC-RV267.
XX
KW Mycobacteriosis; survival; virulence; protective antigen; vaccine;
KW Mycobacterial disease; tuberculosis; leprosy; ds; cosmid.
XX
OS Mycobacterium tuberculosis.
XX
FN WO200274903-A2.
XX

PD 26-SEP-2002.
XX
PF 22-FEB-2002; 2002WO-IB001973.
XX
PR 22-FEB-2001; 2001US-0270123P.
XX
PA (INSP) INST PASTEUR.
XX
PI Cole S;
XX
DR WPI; 2002-759885/82.
XX
PT Identifying and selecting genes for survival or virulence of mycobacteria by a comparative genomic analysis of the sequences of Mycobacterium tuberculosis and *M. leprae*.
XX
PS Disclosure; Fig 8; 874pp; English.

This invention relates to a novel method for identifying essential genes for survival or virulence of mycobacteria species. The method comprises aligning the genomic sequence of a first mycobacterium species on a genomic sequence of a second mycobacterium species and selecting a polynucleotide sequence that is highly conserved in both genomes with no counterparts in other bacterial genomic sequences and that corresponds to an essential gene for the survival or virulence of mycobacterium species. The method of the invention is useful for detecting *M. tuberculosis* or *M. leprae* infection. The method reduces the number of potential new targets and protective antigens for new drugs and vaccine compositions to treat and prevent mycobacterial diseases, particularly tuberculosis and leprosy. The present sequence represents a Mycobacterium cosmid DNA sequence used in the method of the invention

Query Match 12.6%; Score 40.4; DB 6; Length 45190;
Best Local Similarity 50.4%; Pred. No. 2;
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;

QY 72 TGGGGGCAACATGTTGGCGCGCGAGCTTCGGTCCGCGGAGAGCGCGCGCCCAACG 131
DB 4805 TCGGCTGCAACAGACCGTGGCGGAGCACTCGCACAGCGAGTTACCGCACTAACCGTGGT 4864
QY 132 GCAGCGCGACCTCGAGGCTATCCCGCGAGCGTA-----CGCCGAGATGGTGGGACATC 185
DB 4865 CAATCGCGACGCTCTGGCCCTGCGCGGAGAGATCTAGCGCGGCGCGCGCGTGGT 4924
QY 186 ACAGAAATCGACAGACGACGCGGAGTTCGGTTCGATCGATCTCGACGATGA 245
DB 4925 TGCCAATCTGCGGTACACGATAGCGGTACCGCGGTTGTCATCTGTCGAGTTCCC 4984
QY 246 GCAGCGCGGCTTCAGCGGAGTGAAGCTCGGCTGGCATCCGTATCACTTCCCGCGAGA 305
DB 4985 GTCGATCGGTGTCGTCGCGGTGATGGTCAGCGCGAGGTGCGCGAACCGCTCGCGCGGA 5044
QY 306 ACCGACAGCAACAG 321
DB 5045 GCCGGGCAACAGAG 5060

RESULT 15
AAI99682.11
Continuation (12 of 45) of AAI99682 from base 1100001 (Mycobacterium tuberculosis strain WP Sequence split into 45 fragments LOCUS AAI99682 Accession AAI99682
WP Fragment Name Begin End
WP AAI99682_00 1 110000
WP AAI99682_01 100001 210000
WP AAI99682_02 200001 310000
WP AAI99682_03 300001 410000
WP AAI99682_04 400001 510000
WP AAI99682_05 500001 610000
WP AAI99682_06 600001 710000
WP AAI99682_07 700001 810000
WP AAI99682_08 800001 910000

WP	AAI99682_09	900001	1010000
WP	AAI99682_10	1000001	1110000
WP	AAI99682_11	1100001	1210000
WP	AAI99682_12	1200001	1310000
WP	AAI99682_13	1300001	1410000
WP	AAI99682_14	1400001	1510000
WP	AAI99682_15	1500001	1610000
WP	AAI99682_16	1600001	1710000
WP	AAI99682_17	1700001	1810000
WP	AAI99682_18	1800001	1910000
WP	AAI99682_19	1900001	2010000
WP	AAI99682_20	2000001	2110000
WP	AAI99682_21	2100001	2210000
WP	AAI99682_22	2200001	2310000
WP	AAI99682_23	2300001	2410000
WP	AAI99682_24	2400001	2510000
WP	AAI99682_25	2500001	2610000
WP	AAI99682_26	2600001	2710000
WP	AAI99682_27	2700001	2810000
WP	AAI99682_28	2800001	2910000
WP	AAI99682_29	2900001	3010000
WP	AAI99682_30	3000001	3110000
WP	AAI99682_31	3100001	3210000
WP	AAI99682_32	3200001	3310000
WP	AAI99682_33	3300001	3410000
WP	AAI99682_34	3400001	3510000
WP	AAI99682_35	3500001	3610000
WP	AAI99682_36	3600001	3710000
WP	AAI99682_37	3700001	3810000
WP	AAI99682_38	3800001	3910000
WP	AAI99682_39	3900001	4010000
WP	AAI99682_40	4000001	4110000
WP	AAI99682_41	4100001	4210000
WP	AAI99682_42	4200001	4310000
WP	AAI99682_43	4300001	4410000
WP	AAI99682_44	4400001	4411529
Query Match 12.6%; Score 40.4; DB 4; Length 110000;			
Best Local Similarity 50.4%; Pred. No. 2.2;			
Matches 129; Conservative 0; Mismatches 121; Indels 6; Gaps 1;			
OY	72	TGGGGCGCAACATGGTTGGCGCGCCGACGTTCCGTCGCGCGAGCAGCGCGCGCCCAACG	131
Db	29425	TCGGCTGCAACAGACCGGTGGCGGAGCACTCGCACAGCGAGGTTACCGACTAACGGTGGT	29484
OY	132	GCAGCGGACCTCGAGCTATCGCCGAGCGTA-----CGCCGAGATGGTGGCGACATC	185
Db	29485	CAATCGGAGCTCTGGCCCTGCGCGGAGGATCTAGCCGCGCGCCGACCGCGTGGT	29544
OY	186	ACACGAAATCGACGACGACACACCGCAACTGGCGGTGTTGTCGATCATCTCGACGATGA	245
Db	29545	TGCCAATCTGCCGTACAACGTAGCGGTACCGGGGTTGTTGCATCTGCTTGTGAGTTCCC	29604
OY	246	GCAGCGCGCGCTTGGCGGGGATGAAGTCGCTGGCATCCGTATCACTTCCCGACGA	305
Db	29605	GTGATCCCGTGTGTCACCGGTGATGTGAGCCGCGAGGTGCGCGAACGGCTCGCCCGCA	29664
OY	306	ACCCGACAGCAACAG	321
Db	29665	GCCGGCGAGCAAGAG	29680

Search completed: February 10, 2005, 05:34:17
Job time : 346 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 10, 2005, 03:45:06 ; Search time 2015 Seconds
(without alignments)
7533.496 Million cell updates/sec

Title: US-10-647-089-93

Perfect score: 321

Sequence: 1 ttgaccacacgacgactaa.....acgaaccgacagcaaacag 321

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 4526729 seqs, 23644849745 residues

Total number of hits satisfying chosen parameters: 9053458

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	321	100.0	321	6	AR456084
2	321	100.0	324	6	AX074032 Sequence
3	321	100.0	110000	1	Continuation (30 o
4	321	100.0	346051	1	EX842580 Mycobacte
5	309	96.3	42061	1	MSGY223
6	47.2	14.7	125020	9	AF429315
7	46.2	14.4	64508	1	UBA564903
8	46.2	14.4	67066	1	AB063332
9	46.2	14.4	87888	1	AY365053
10	46.2	14.4	108845	1	U66917
11	45.4	14.1	301457	1	AB016924
12	44.8	14.0	320150	1	AP005033
13	44.4	13.8	12201	1	AE012255
14	44.2	13.8	104289	8	CNS08C7V
15	44.2	13.8	110000	2	LMFLCHR34_10
16	43.8	13.6	13121	1	AE012257
17	43.6	13.6	303550	1	SC0939118
18	43	13.4	2832	8	AY452532
19	42.8	13.3	13486	1	AE004842

20	42.8	13.3	71286	1	AF143772
21	42.8	13.3	301068	1	AE017231
22	42.6	13.3	131946	8	AC129226
23	42.4	13.2	646	8	AK062705
24	42.4	13.2	168560	8	AC118132
25	42	13.1	299750	1	AP005964
26	42	13.1	299925	1	AP005039
27	41.8	13.0	106873	14	AB096160
28	41.8	13.0	147891	8	AP004693
29	41.8	13.0	163371	8	AP004585
30	41.6	13.0	10507	1	AE005681
31	41.4	12.9	216050	1	AL646076
32	41.2	12.8	289308	1	AE017242
33	41.2	12.8	309267	1	AE017235
34	41	12.8	400	6	CQ462190
35	41	12.8	10772	1	AE011659
36	40.8	12.7	299050	1	SC0939104
37	40.8	12.7	300450	1	AP005960
38	40.8	12.7	300917	1	AE016919
39	40.8	12.7	349260	1	EX572595
40	40.6	12.6	4161	8	AK072093
41	40.6	12.6	86896	1	RCU57682
42	40.6	12.6	151269	8	AC016781
43	40.6	12.6	160365	2	AC084764
44	40.6	12.6	299425	1	AP005037
45	40.6	12.6	314100	1	SC0939106

ALIGNMENTS

RESULT 1	AR456084	Sequence 93 from patent US 6686166.	321 bp	DNA	linear	PAT 20-FEB-2004
LOCUS	AR456084	Sequence 93 from patent US 6686166.	321 bp	DNA	linear	PAT 20-FEB-2004
DEFINITION	AR456084	Sequence 93 from patent US 6686166.	321 bp	DNA	linear	PAT 20-FEB-2004
ACCESSION	AR456084.1	GI:42691080				
VERSION	AR456084.1	GI:42691080				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1 (bases 1 to 321)					
AUTHORS	Behr, M., Small, P., Schoolnik, G. and Wilson, M.A.					
TITLE	Molecular differences between species of the M. tuberculosis complex					
JOURNAL	Patent: US 6686166-A 93 03-FEB-2004;					
FEATURES	Location/Qualifiers					
source	1..321					
ORIGIN	/organism="unknown"					
	/mol_type="genomic DNA"					

Query Match	100.0%;	Score 321;	DB 6;	Length 321;
Best Local Similarity	100.0%;	Pred. No. 5.9e-49;		
Matches 321;	Conservative	0;	Mismatches	0;
			Indels	0;
			Gaps	0;
Qy	1	TTGACCCCAAGCGCCTAAAGCCGACGAGGCTCGCCGAGGCTCAAGCCCGCGT	60	
Db	1	TTGACCCCAAGCGCCTAAAGCCGACGAGGCTCGCCGAGGCTCAAGCCCGCGT	60	
Qy	61	CGGAATCGGTTGGCGCGCAACATGGTGGCGCGGCGGCTTCGTCGCGCAGAGCGC	120	
Db	61	CGGAATCGGTTGGCGCGCAACATGGTGGCGCGGCGGCTTCGTCGCGCAGAGCGC	120	
Qy	121	CGCGCCCAAGCGCAGGCTCGAGGCTATCCGCGGAGGCTAGCCGAGATGGTGGCG	180	
Db	121	CGCGCCCAAGCGCAGGCTCGAGGCTATCCGCGGAGGCTAGCCGAGATGGTGGCG	180	
Qy	181	ACATCACAAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACG	240	
Db	181	ACATCACAAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACG	240	
Qy	241	GATGACGACGCGCGGCTTTGAGCGGGGATGAAGTCGGCTGGCATCCGCTATCATTCC	300	

Db 241 GATGAGCAGCGCGCTTGGCGGGGATGAAGCTCGCTGGCATCGTATCACTTCCCC 300
QY 301 GACGAACCCGACGCAACAG 321
Db 301 GACGAACCCGACGCAACAG 321

RESULT 2

AX074032
LOCUS AX074032 324 bp DNA linear PAT 06-FEB-2001
DEFINITION Sequence 14 from Patent WO0104151.
ACCESSION AX074032
VERSION AX074032.1 GI:12710257

KEYWORDS

SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.

REFERENCE

1 Andersen, P. and Skjoet, R.
Tuberculosis vaccine and diagnostics based on the mycobacterium tu
berculosis eat-6 gene family
Patent: WO 0104151-A 14 18-JAN-2001;

JOURNAL

Statens Serum Institut (DK)

FEATURES

source Location/Qualifiers
1..324
/organism="Mycobacterium tuberculosis"
/mol_type="unassigned DNA"
/db_xref="taxon:1773"

CDS

1..324
/notes="unnamed protein product"
/codon_start=1
/transl_table=1
/protein_id="CAC28442.1"
/db_xref="GI:12710258"
/translation="MTHKTKRQPAIAAGLNAPRRNRVGRHQHGWADVPASBQRAQR
ORDLEAIRRAYAEWATSHIEDDTAELLMSHLIDDEQRLEAGMKLGWHFYPFDE
PDSKQ"

ORIGIN

Query Match 100.0%; Score 321; DB 6; Length 324;
Best Local Similarity 100.0%; Pred. No. 5.9e-49;
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGACCCACAGCGCACTAAACGCCAGCGCATCGCGCGGCGTCAACGCCCGCGT 60
Db 1 TTGACCCACAGCGCACTAAACGCCAGCGCATCGCGCGGCGTCAACGCCCGCGT 60

QY 61 CGGAATCGGTTGGCGGCAACATGTTGGCGCGCGCATGTTCCGTCGCCCGAGCAGCGC 120
Db 61 CGGAATCGGTTGGCGGCAACATGTTGGCGCGCGCATGTTCCGTCGCCCGAGCGCGC 120

QY 121 CGCGCCCAACGCGGCGGCGCTCGAGGCTATCCGCGGCGGCTACGCCGAGATGTTGGCG 180
Db 121 CGCGCCCAACGCGGCGGCGGCTCGAGGCTATCCGCGGCGGCTACGCCGAGATGTTGGCG 180

QY 181 ACATCACAGAAATCCAGCAGCACAGCGGCTGTTGGCGGCTGTTGCGATGATCTCGAC 240
Db 181 ACATCACAGAAATCCAGCAGCACAGCGGCTGTTGGCGGCTGTTGCGATGATCTCGAC 240

QY 241 GATGAGCAGCGCGGCTTGGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 300
Db 241 GATGAGCAGCGCGGCTTGGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 300

QY

301 GACGAACCCGACGCAACAG 321

Db

301 GACGAACCCGACGCAACAG 321

RESULT 3

AE000516_29/c

WPCOMMENT

Sequence split into 44 fragments LOCUS AE000516 Accession AE000516

Fragment Name	Begin	End
AE000516_00	1	110000
AE000516_01	100001	210000
AE000516_02	200001	310000
AE000516_03	300001	410000
AE000516_04	400001	510000
AE000516_05	500001	610000
AE000516_06	600001	710000
AE000516_07	700001	810000
AE000516_08	800001	910000
AE000516_09	900001	1010000
AE000516_10	1000001	1110000
AE000516_11	1100001	1210000
AE000516_12	1200001	1310000
AE000516_13	1300001	1410000
AE000516_14	1400001	1510000
AE000516_15	1500001	1610000
AE000516_16	1600001	1710000
AE000516_17	1700001	1810000
AE000516_18	1800001	1910000
AE000516_19	1900001	2010000
AE000516_20	2000001	2110000
AE000516_21	2100001	2210000
AE000516_22	2200001	2310000
AE000516_23	2300001	2410000
AE000516_24	2400001	2510000
AE000516_25	2500001	2610000
AE000516_26	2600001	2710000
AE000516_27	2700001	2810000
AE000516_28	2800001	2910000
AE000516_29	2900001	3010000
AE000516_30	3000001	3110000
AE000516_31	3100001	3210000
AE000516_32	3200001	3310000
AE000516_33	3300001	3410000
AE000516_34	3400001	3510000
AE000516_35	3500001	3610000
AE000516_36	3600001	3710000
AE000516_37	3700001	3810000
AE000516_38	3800001	3910000
AE000516_39	3900001	4010000
AE000516_40	4000001	4110000
AE000516_41	4100001	4210000
AE000516_42	4200001	4310000
AE000516_43	4300001	4403837

Continuation (30 of 44) of AE000516 from base 2900001 (AE000516 Mycobacterium tuberculosis)

Query Match	Score 321; DB 1; Length 110000;
Best Local Similarity 100.0%; Pred. No. 2.2e-49;	
Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;	
QY 1 TTGACCCACAGCGCACTAAACGCCAGCGCATCGCGCGGCGTCAACGCCCGCGT 60	
Db 71695 TTGACCCACAGCGCACTAAACGCCAGCGCATCGCGCGGCGTCAACGCCCGCGT 71636	
QY 61 CGGAATCGGTTGGCGGCGCAACATGTTGGCGGCGGCGTTCGTCGCCCGAGCAGCGC 120	
Db 71635 CGGAATCGGTTGGCGGCGCAACATGTTGGCGGCGGCGTTCGTCGCCCGAGCAGCGC 71576	
QY 121 CGCGCCCAACGCCAGCGGCGACCTCGAGGCTATCCGCCGAGCGTAGCCGAGATGTTGGCG 180	
Db 71575 CGCGCCCAACGCCAGCGGCGACCTCGAGGCTATCCGCCGAGCGTAGCCGAGATGTTGGCG 71516	
QY 181 ACATCACAGAAATCCAGCAGCACAGCGGCTGTTGTCGATGATCTCGAC 240	
Db 71515 ACATCACAGAAATCCAGCAGCACAGCGGCTGTTGTCGATGATCTCGAC 71456	
QY 241 GATGAGCAGCGCGGCTTGGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 300	
Db 71455 GATGAGCAGCGCGGCTTGGCGGGGATGAAGCTCGGCTGGCATCCGTTATCACTTCCCC 71396	


```
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAA16049.1"
/db_xref="GI:2791512"
/translation="WMKRIAVRLPGEVITFDSEVSOIRIPSRRAAVVLRASNASDAA
ILTATPNHLDLALGAQAALPTSIDAAHPARPARDPCLYPTGQALPRTG"
gene
CDS
2532..3248
/locus_tag="Rv2473"
/locus_tag="Rv2473"
/functions="UNKNOWN"
/notes="Rv2473, (MTV008.29), len: 238 aa. Possible
pro-,ala-rich membrane protein, with possible
transmembrane domain around aa 81-104."
/codon_start=1
/transl_table=11
/product="POSSIBLE ALANINE AND PROLINE RICH MEMBRANE
PROTEIN"
/protein_id="CAA16050.1"
/db_xref="GI:2791513"
/translation="MAPTSSVASSELLMPFSAAGVVGWRTTATASORYHRPMSDT
PFAEPYDQRPQPGVPPGWDGSSRPAPSTTPSPGRWSILVALALAVSLGVIGW
FHPQHDKSPASATFTSQISDAKENVCAAHRIVRQAVLNTQNPVPDPTGD
LAVANARLALYSGDYLLRLRTAEPAETPAELRDAVRSLANALQELAVNYLAGPDSV
VPLRLALERTDRAVDPLCV"
gene
CDS
3280..3933
/locus_tag="Rv2474c"
/locus_tag="Rv2474c"
/functions="UNKNOWN"
/notes="Rv2474c, (MTV008.30c), len: 217 aa. Hypothetical
protein. Shows weak similarity with Q9L246|SCSD10.18c
HYPOTHETICAL 24.9 KDA PROTEIN from Streptomyces coelicolor
(238 aa), FASTA scores: opt: 111, E(): 5.6, (30% identity
in 233 aa overlap), BLASTP scores: Score= 135, E= 3.5e-07,
P= 3.5e-07, Identities= 55/182 (30%)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/protein_id="CAA16051.1"
/db_xref="GI:2791514"
/translation="WVERGLWLPDPAHRADLAFVDFHALRLDDAAVIRARSTGLLS
AWATGFDVLASRVAGKVRPDDLVAARSALHGLATTDASGYDFGYSDMSAWRGL
PPESGTYLDDVDFARVLMDLAHRGALKEHSSAGPVPVSLLDQEVIVQVSSADVVVGL
PNECVFALTAMGFLPQSALTSADLIRISPAWLRDLARFSGVYVRRHGAALVLR"
gene
CDS
3939..4355
/locus_tag="Rv2475c"
/locus_tag="Rv2475c"
/functions="UNKNOWN"
/notes="Rv2475c, (MTV008.31c), len: 138 aa. Conserved
hypothetical protein, showing similarity with
Q9L245|SC6D10.19c HYPOTHETICAL 16.2 KDA PROTEIN from
Streptomyces coelicolor (136 aa), FASTA scores: opt: 236,
E(): 1.9e-09, (34.1% identity in 126 aa overlap). Also
some similarity with AAK44393|Z97050|MTC128_3 CONSERVED
HYPOTHETICAL PROTEIN from Mycobacterium tuberculosis
cosmid 1 (151 aa), FASTA scores: opt: 147, E(): 0.00025,
(29.2% identity in 120 aa overlap)."
/codon_start=1
/transl_table=11
/product="CONSERVED HYPOTHETICAL PROTEIN"
/Query Match
/Best Local Similarity 100.0%; Score 321; DB 1; Length 346051;
/Matches 321; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 TTGACCCACGAGCGCACTAAACCCGACGAGCCATCGCGCAGGGCTCAACGCCCGCGGT 60
Db 203873 TTGACCCACGAGCGCACTAAACCCGACGAGCCATCGCGCAGGGCTCAACGCCCGCGGT 203814
QY 61 CGGAATCGCGTTCGGCGGCAACATGTTGCGCGGCGAGCTTCGTCGCCGAGCAGCGC 120
Db 203813 CGGAATCGCGTTCGGCGGCAACATGTTGCGCGGCGAGCTTCGTCGCCGAGCAGCGC 203754
```

```
QY 121 CGGCCCCAACGCGCAGCGGACCTCGAGGCTATCCGCGGACGCTACGCCGAGATGTTGGCG 180
Db 203753 CGGCCCCAACGCGCAGCGGACCTCGAGGCTATCCGCGGACGCTACGCCGAGATGTTGGCG 203694
QY 181 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 240
Db 203693 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 203634
QY 241 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTTCCCC 300
Db 203633 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTTCCCC 203574
QY 301 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 321
Db 203573 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 203553

RESULT 5
MSGY223
LOCUS MSGY223 42061 bp DNA linear BCT 10-DEC-1996
DEFINITION Mycobacterium tuberculosis sequence from clone y223.
ACCESSION AD000019
VERSION AD000019.1 GI:1717737
KEYWORDS
SOURCE Mycobacterium tuberculosis
ORGANISM Mycobacterium tuberculosis
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
Corynebacterineae; Mycobacteriaceae; Mycobacterium; Mycobacterium
tuberculosis complex.
REFERENCE 1 (bases 1 to 42061)
AUTHORS Du,L.
TITLE Direct Submission
JOURNAL Submitted (11-OCT-1996) L.Du, Genome Therapeutics Corporation, 100
Beaver Street, Waltham,-MA, USA, 02154 du@cric.com
COMMENT GSDB:S:1004712.
FEATURES
source
1..42061
/organism="Mycobacterium tuberculosis"
/mol_type="genomic DNA"
/db_xref="taxon:1773"
/clone="y223"
ORIGIN
Query Match 96.3%; Score 309; DB 1; Length 42061;
Best Local Similarity 99.7%; Pred. No. 4e-47; Indels 1; Gaps 1;
Matches 320; Conservative 0; Mismatches 0;
QY 1 TTGACCCACGAGCGCACTAAACCCGACGAGCCATCGCGCAGGGCTCAACGCCCGCGGT 60
Db 5783 TTGACCCACGAGCGCACTAAACCCGACGAGCCATCGCGCAGGGCTCAACGCCCGCGGT 5842
QY 61 CGGAATCGCGTTCGGCGGCAACATGTTGCGCGGCGAGCTTCGTCGCCGAGCAGCGC 120
Db 5843 CGGAATCGCGTTCGGCGGCAACATGTTGCGCGGCGAGCTTCGTCGCCGAGCAGCGC 5902
QY 121 CGGCCCCAACGCGCAGCGGACCTCGAGGCTATCCGCGGCGCTACGCCGAGATGTTGGCG 180
Db 5903 CGGCCCCAACGCGCAGCGGACCTCGAGGCTATCCGCGGCGCTACGCCGAGATGTTGGCG 5962
QY 181 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 240
Db 5963 ACATCACAGAAATCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6022
QY 241 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTTCCCC 300
Db 6023 GATGACGACGCGCGGCTTGAGCGGGGATGAAGCTCGGCTGGCATCCGTATCACTTTCCCC 203574
QY 301 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 321
Db 6082 GACGACCCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 6102
```

```
RESULT 6
AF429315/c      125020 bp      DNA      linear      PRI 18-JAN-2002
LOCUS      Homo sapiens junctophilin 3 (JPH3) gene, partial cds.
DEFINITION
ACCESSION      AF429315
VERSION      AF429315.1 GI:17646244
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 125020)
Holmes,S.E., O'Hearn,E., Rosenblatt,A., Callahan,C., Hwang,H.S.,
Ingersoll-Ashworth,R.G., Fleisher,A., Stevanin,G., Brice,A.,
Potter,N.T., Ross,C.A. and Margolis,R.L.
A repeat expansion in the gene encoding junctophilin-3 is
associated with Huntington disease-like 2
Nat. Genet. 29 (4), 377-378 (2001)
11694876
2 (bases 1 to 125020)
Holmes,S.E., Ingersoll-Ashworth,R.G., Ross,C.A. and Margolis,R.L.
Direct Submission
Instituted (05-OCT-2001) Psychiatry, Johns Hopkins Medical
Institutions, 600 N. Wolfe St., Baltimore, MD 21287, USA
JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 125020
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q24.3; between D16S520 and WI-12410"
/note="Isolated from a patient with Huntington's
Disease-Like 2 (HDL2)."
Complement(35501..35746)
/rpt_unit="ctg"
/rpt_type="tandem"
complement(<36507..>36887)
/gene="JPH3"
/note="synonym: JP3"
complement(<36507..>36887)
/gene="JPH3"
/product="junctophilin 3"
complement(<36507..>36887)
/gene="JPH3"
/note="component of the junctional complex between plasma
membrane and endoplasmic reticulum"
/codon_start=1
/product="junctophilin 3"
/protein_id="AAL40941.1"
/db_xref="GI:17646245"
/translation="MSSSGRFNFDGSGYCGWEDGKAHGVCVTPKPGCGEYTCSSWS
HGFEVIGYVTPSGNYSQVTAQGRKHGIGLESKGVYKGVWTHGFKRGYGVRECA
NGAKYEGTWSNGLOPGYGTETYSVG"
ORIGIN
Query Match      14.7%; Score 47.2; DB 9; Length 125020;
Best Local Similarity 10.6%; Pred.No.19;
Matches 31; Conservative 149; Mismatches 112; Indels 0; Gaps 0;
QY      20 AACGCAGCCAGCCATCGCCGAGGCTCAACGCCCGCGTCGGAATCGCTGGCGCGC 79
D      :|: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      51134 MACMYWYKSRRTKSMATSGCWRMGAMESKGMRYKWSCRYKGMRWGWSYKCSGYM 51075
QY      80 AACATGTTGGCGCGCGCGAGTTCGTCGCGCAGCAGCGCGCCGACCGGAGCGCG 139
D      :|: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      51074 SMGMYSGRSTSYCYSTGTCYKCAKMKCKSKTSKSCCTKSKYSCSMGRYSYCTCWGG 51015
QY      140 ACCTCAGAGCTATCGCCGAGGTACGCGAGATGCTGGCGACATCACACGAAATCGAGC 199
D      :|: : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      51014 RMCKGMKSSCCASRSRGSYKRWAGSWRCRGRGRGMSCWYSSWWSKYSMASYKSSARGMK 50955
QY      200 ACGACACAGCCGAACTGGCGCTGTTGTCATGATCATCTCGACGATGAGCAGCGCGCTTG 259
```

```
Db      50954 CYKTSKMYYSMBKSHDBSGYKYSRYKSYBYWTMMKDSMCVSDSHMMYDMYDKMR 50895
QY      260 AGGCGGGGATGAAGCTCGGCTGGCATCGGTATCATCTCCCGCAGCAACCCGA 311
Db      50894 MKVSSGTRGHBHSAWMMCKMSRSDSGYSGMYHMKSYWKSASAKHMYISGM 50843

RESULT 7
UBA564903/c      64508 bp      DNA      circular BCT 22-JAN-2004
LOCUS      Uncultured bacterium Incp-ibeta multiresistance plasmid pB10.
ACCESSION      AJ564903
VERSION      AJ564903.1 GI:31746361
KEYWORDS      'tnpA gene; acetyltransferase;
aminoglycoside-3',-phosphotransferase;
aminoglycoside-6-phosphotransferase; ATPase, autophosphorylase;
beta-lactamase precursor; coregulator protein; damage
inducible-like protein; dihydropteroate synthetase; din gene; DNA
integrase; DNA relaxase; DNA replication primase; DNA
topoisomerase; DNA transport protein; hypothetical transmembrane
protein; hypothetical Xf2080 protein; incC1 gene; incC2 gene; incC2
protein; inclusion membrane protein; integral membrane protein for
mercuric transport; int11 gene; kfrA gene; KfrA protein; klcA gene;
KlcA protein; klcB gene; KlcB protein; klcE gene; KlcE protein; klcA
protein; klcE gene; KlcE protein; klcF gene; KlcF protein; klcA gene; KlcA
protein; korB gene; KorB protein; korC gene; KorC protein; KorA
maturation peptidase; merA gene; mercuric reductase; mercury
resistance protein; merD gene; merE gene; merP gene; merR gene;
merT gene; muramidase; ORF-2; ORF5; ORFE-like; origin of transfer;
oriT; oriT binding protein; outer membrane protein; oxa-2 gene;
pecM-like gene; periplasmic mercury ion binding protein; prepilin;
qacEdelta1 gene; relaxosome stabilisation protein; replication
initiation protein; repressor protein; resolvase; single-strand DNA
binding protein; small exporter protein; sbb gene; sra gene; strB
gene; sul1 gene; teta gene; tetR gene; tetracycline efflux protein;
tetracycline resistance repressor protein; tnpA gene; tnpA' gene;
tnpR gene; trac gene; traB gene; traI gene; traJ gene; traK gene; traL
gene; traM gene; TraM protein; traN gene; traO protein; traP gene;
origin protein; traQ gene; traR gene; trbD protein; trbE gene; trbF
protein; trbF gene; trbF protein; trbG gene; trbG protein; trbH
gene; trbH protein; trbI gene; trbI protein; trbJ gene; trbJ
protein; trbK gene; trbK protein; trbL gene; trbM gene; trbM
protein; trbN gene; trbO gene; trbO protein; trbP gene; trbP
protein; trfai gene; truncated transposase; upf30.5 gene; xif2080
gene.
SOURCE      uncultured bacterium
ORGANISM      uncultured bacterium
Bacteria; environmental samples.
REFERENCE
1 Schluter,A., Heuer,H., Szczepanowski,R., Forney,L.J., Thomas,C.M.,
Puhler,A. and Top,E.M.
The 64 508 bp IncP-ibeta antibiotic multiresistance plasmid pB10
isolated from a waste-water treatment plant provides evidence for
recombination between members of different branches of the
Incp-ibeta group
Microbiology (Reading, Engl.) 149 (Pt 11), 3139-3153 (2003)
JOURNAL      22962753
MEDLINE      14600226
PUBMED
REFERENCE
2 (bases 1 to 64508)
Schluter,A.
Direct Submission
AUTHORS
TITLE
JOURNAL
FEATURES
source
1. 64508
/organism="uncultured bacterium"
/mol_type="genomic DNA"
/isolation_source="activated sludge"
/db_xref="taxon:77133"
```

```

/environmental_sample
/lab_host="Pseudomonas sp. B13"
/plasmid="pB10"
97. .459
/gene="trbA"
/gene="trbA"
97. .459
/function="regulatory protein for mating pair formation"
/codon_start=1
/transl_table=11
/product="trbA protein"
/protein_id="CAD97487.1"
/db_xref="GI:31746362"
/db_xref="TrEMBL:Q7X3F3"
/translations="MNYIPFTNVRLRLLDERHMTKKELSDRSQVSIISFLSLLTTGKAN
PSLKVMEIDIAQALETLPILLESTDLKDALDLAGGKAPRSIPPGFERVAALVPEHQ
AFIVKKGAEATKKLRGS"
769. .1731
/gene="trbB"
769. .1731
/gene="trbB"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="putative ATPase, autophosphorylase"
/protein_id="CAD97488.1"
/db_xref="GI:31746363"
/db_xref="TrEMBL:Q7X3F2"
/translations="WTDQPEHNSIKERAKKKLERDMGPELLAALNDPKTVEIMLNAD
GKMLERLEGPMCKIGLRQAQAIETIAGVHGKEVTRSKPILSGELPLDGSFAG
QLPVPVPAFTFAIKKAAFTLDQYVERGIMTAPQREALIAAFAHNNILVIGTGS
GKTLTVAINENVIQDPTEVFIIIEDTGEIOCAENYVQYHTSIDNNMTALKITLIR
MRPRLIVGVRGEALDILMAWNTQHEGGAATHANNAKAGLDRLAMLISMHPDSPK
PIRPLIGEAVHVVHIAIVEGSRRIQEIILEVSFGANGYITKTL"
1748. .2212
/gene="trbC"
1748. .2212
/gene="trbC"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="putative prepilin"
/protein_id="CAD97489.1"
/db_xref="GI:31746364"
/db_xref="TrEMBL:Q7X3F1"
/translations="MQATPAPRLNRNALFYMLGFALLAPFLAPOHAFASBGTGSSL
PYBSWLTNLRNSVTGPVAFALSTIGIVVAGGILIFGELNGFFRTLIFVLWALLVG
AONMSTFFGREGIEAALTDGALHQVKVAALDVAGPFGAAMWAERHGIAG"
2216. .2527
/gene="trbD"
2216. .2527
/gene="trbD"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbD protein"
/protein_id="CAD97490.1"
/db_xref="GI:31746365"
/db_xref="TrEMBL:Q7X3F0"
/translations="MALRAIPIRAGNRNENLFWGGDRELVMFSGLLAGALIFSAQEV
ATVFGALWFGALFALRLMAKADPKURHVILRHRRYKPYIPARSTFFRDNTPSQKQY
K"
2524. .5082
/gene="trbE"
2524. .5082
/gene="trbE"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbE protein"
/protein_id="CAD97491.1"
/db_xref="GI:31746366"
/db_xref="TrEMBL:Q7X389"

```

```

/translation="MIEAIAIAIVLGVALLLILFARIVRAVDAELKLLKHKRAKDAGLA
DLINVAADVDDGVIKNGSGFAAMLYKGDNDASSTEEOREMVSFRINQALAGLNGW
MYHDAVRPAPNYSERGVSFSDSVSAIDERRRRLRGLGAMGEGYFVLVTTWPPF
VLAQKRFELMFDDDAVADPHARTITGLIAQKREITSLERSLSAVATWRJURQKVV
SEEGREVTDDFLSWLQFCVGLHPHVMPLPSNPMDLDAVIGGOELWGGVVPKIGKFI
QVVAIEGPLESTPGVLASALBEPSEYRWSSRFIFMDQHESLKHLDFKRWKQKIRG
PFDQVNTNGTSINODAAWVGDAEAAAEVNSGLVAAGYTTSVVLMDEDRERLAAS
ALLKAVNRLAFAARIETINTDLAVLGLSPGHGVNVRPLINTMNADLLPTSSIM
TGSATPCPMYPLSPALMHCVTVGATPRNLNHRVDLGHTEFPTGAGKSTHGLII
AAQLRYAGMSIYADTKGMSYPLAAGIRATKSGKSLHFTVAADDDRLAFPLQPLE
TKGDRAMAWEMIDTILALNGVNTTTPAQNEIGNAIMSHASGARTLSEFSVLIQDEAI
REAIKOYTVDGSMLHDLDAEBGLSDFTVEIBEMLNIGKFLALPVLILYFRIER
ALKQSPSVIILDEAWMLHCHPAFRAKIREWLKRLKANCLATVMAQSLDAANSGLID
VIVGSTATKIPLPNVVARDETSALYRRMGLNAROIEILATAI PKROYVYVSENGRRL
YDLALGPALATVGSADKESVATIKSLEAKFGHEVHWEWLAGRLNLDYLEA"
5079. .5861
/gene="trbF"
5079. .5861
/gene="trbF"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbF protein"
/protein_id="CAD97492.1"
/db_xref="GI:31746367"
/db_xref="TrEMBL:Q7X3E8"
/translations="MSIADTLKGLIFKPPANSAEPRDRSRKEPLAGGRAGEENPY
LAARTWHDHGVSVVSQRTQWQVIGILSLIALAGVGVHIGSGSKETPYVVEVDKL
TAAAGVQVQAKPADPRVTHAAVADMSWCARMSPDVALQKRCVKFAYSLMADNDPA
TFKMEWNLNGPDSPPFRAEKEMVSVIEIKTVPQTPDTQVQWVETTRDROGTLKQ
PYTWALVYTTIAEVTNTTDEQLNNPLSIYVSDYSWRIQ"
5858. .6778
/gene="trbG"
5858. .6778
/gene="trbG"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbG protein"
/protein_id="CAD97493.1"
/db_xref="GI:31746368"
/db_xref="TrEMBL:Q7X3E7"
/translations="MNSRSTMMKTLTSLIVGAAVAVPSLAAAPGDDIADKYFNKTN
PTLTAQERAAALAIKRWAGSAGATIGKPVAGSGAIKFVYGAQQPIVCAVLQCVQVAL
QPEOVNINIGLDTARWTEPAITGSGATEVQHLLIKPMVDGLTSLVTTNRRSVHF
RLRSRTYMPQVSPFTYPEAQAKWDATQREKQERTNTLTPATGVLGDLSDYELS
GSASKPVRVNDGKTIIEPCTMQOOTEAPTLIVRKEGGLFTDDETVLVNRYVQGN
RIIVDTVFDRLIAGVSSQDRVTISRRK"
6781. .7269
/gene="trbH"
6781. .7269
/gene="trbH"
/function="involved in mating pair formation"
/codon_start=1
/transl_table=11
/product="trbH protein"
/protein_id="CAD97494.1"
/db_xref="GI:31746369"
/db_xref="TrEMBL:Q7X3E6"
/translations="MRKIVSLALLALGLGCGCATTSYGNFVQSAALDQOKLATDAV
QOLATLYAPARTLRLQOPTPDPPGQALVLSLRDKGYALLLEYAPAGATQAPASAQ
ASAPATTAAAGGLPLRYVLDQAGDNLRLTMVGNQSIITRPPYLAQNGTFAPAGYVVR
KE"
7274. .8695

```

```

Query Match      14.4%; Score 46.2; DB 1; Length 64508;
Best Local Similarity 47.7%; Pred. No. 33;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

QY      6 CCACAAGCGCACTAAACGCCAGCCAGCCATCGCCGCGGCTCAACGCGCCCGCGTCGGAA 65
DB      35080 CGACAAAGCCCGAGGAGAGGGCCACGCCGCCACGAGGTGGAGGCGCTCGGGCCGA 35021

```

Qy 66 TCAGCTTGGCGGCAACATGTTGGCGGCGGCGAGCGTTCGCTCCGCGGAGCAGCGCGCGC 125
Db 35020 CCTCGCGCGCGCAACAGCGGTACCGCGGAGATCGAGCGCGCGCGCGGAGCTCGCTGC 34961
Qy 126 CCAAGCGGAGCGCACCTCGAGGCTATCCGCGGAGCGTACCGCGAGATGGTGGCGACATC 185
Db 34960 CGACTTGGAGCGGCGCAACAGCGGCGGCGGCGGCGCGCGCGCGCGCGCGCGCGCG 34901
Qy 186 ACAGAAATCCAGCAGCAGCAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 245
Db 34900 GAAGCCACCGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 34841
Qy 246 GCAGCGCGCGCTTGAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 288
Db 34840 GCAGCGCAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG 34798

RESULT 8
AB063332/c 67066 bp DNA circular BCT 16-JUL-2004
LOCUS AB063332
DEFINITION Delftia acidovorans plasmid pu01 DNA, complete sequence.
ACCESSION AB063332
VERSION AB063332.1 GI:34013299
KEYWORDS
SOURCE Delftia acidovorans
ORGANISM Delftia acidovorans
Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
Comamonadaceae; Delftia.

REFERENCE
1 Sota, M., Endo, M., Nitta, K., Kawasaki, H. and Tsuda, M.
AUTHORS Characterization of a class II defective transposon carrying two
TITLE haloacetate dehalogenase genes from Delftia acidovorans plasmid
pu01
JOURNAL Appl. Environ. Microbiol. 68 (5), 2307-2315 (2002)
MEDLINE 2197280
PUBMED 11976102

REFERENCE
2 Sota, M., Kawasaki, H. and Tsuda, M.
AUTHORS Structure of haloacetate-catabolic Incp-1beta plasmid pu01 and
TITLE genetic mobility of its residing haloacetate-catabolic transposon
JOURNAL J. Bacteriol. 185 (22), 6741-6745 (2003)
MEDLINE 22957203
PUBMED 14594853

REFERENCE
3 (bases 1 to 67066)
AUTHORS Sota, M., Tsuda, M. and Kawasaki, H.
TITLE Direct Submission
JOURNAL Submitted (19-JUN-2001) Masahiro Sota, Institute for Environmental
Sciences, Department of Environmental Simulation; 1-7 Ienomae,
Obuchi, Rokkasho, Aomori 039-3212, Japan (E-mail: sota@ies.or.jp,
Tel: 81-175-71-1407, Fax: 81-175-71-1401)
PEATURES Location/Qualifiers
1..67066
/organism="Delftia acidovorans"
/mol_type="genomic DNA"
/strain="B"
/db_xref="taxon:80866"
/plasmid="pu01"
/country="Japan:Osaka"
/note="Inc-beta plasmid-the organism was originally
submitted as Moraxella sp. B"
1..901
/note="vegetative origin (oriV)"
897..901
/note="target sequence of TnHad2"
902..18505
/note="Tn3-family, Tn21-group, Tn21-subgroup"
/transposon="TnHad2"
902..939
/rpt_type="inverted"
949..9884
/transposon="TnHad1"
949..4152
/note="IS1071"

repeat_origin
misc_feature
repeat_region
repeat_unit
repeat_region
repeat_region

1094..4009
/gene="tnpA"
1094..4009
/gene="tnpA"
/function="transposition"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="transposase"
/protein_id="BAC81973.1"
/db_xref="GI:34013300"
/translation="MQSHHTTFLGMRGLPRDISDFEMKAFPTFGAERDAINARRGDS
HKLGLALHIGFLRMSGLLGAFRVIVPALMRHLGNELGIAPEVASLRAMTERGRTLF
DHOQVACTVLGFQWMSHQRSLVRELDEVARCADRDQLLVRARQWLYKKNLVIYVHE
RAIRTLIAAALQLEVTGTATASVDPATLDRWRASVSELRPDGTQQSWLWAPAK
HSTROI SEVLERIDLLYLDVHKHLADI POLILRRYARLVSPSPSAGAKI KEPARTV
EVACFLRYCLFTTDDLI MWORRIADLRQAADVPATVNMWAMYKTLGELVALSA
QGAVPDABLRARLEALITETQKRPSPRASLIREGLIDGIRPVRSLLVAIAKLWQAT
GEHPAIEYLAKQALYDKGRKLPVEVVAFPSLQMWISSPSPRERAFQALEVATLIF
ALRRVNRGSGVWIEHSLSPFRARLFTDERWQASKKHYARLSLPSKAATFLKPLLA
RVTAGVDVAAAARSGLVRVDDLHLSPLPAEDDEPVTKLRAALDHRIGEVOLPEVI
LAVDAQVRFSWIMLGRPRSTDELLMYYAGTMAHGTSLTAVECARMIPQASATLROA
MRWARDERLSQAQVLEFMORHPAAATGRSDSLASDDMMSTTKRWQARLDPRR
NTESIGIYSHVKORWIFHAQPFVLRQAGVAVIEGVIROEKLSTSLAVDTHGYTDF
AMSHARILGDFDLPRKLKQRLHVPGRGTPVPAEIAVCEANVDVALIEKHWSLVH
LAASVMSGHASAVAALAFGSAQGDPIYEAGVQLGRLLRTAFADLVFVKDAFNELR
RVLNRGVAVNALKRAIYTGRTISPAQAKRVDEMQAADALSIMANI VMAWNTSOMQAVL
DRWSNRQVTPPELIGKIAPTLEBSINLGRVFRFPVDRYADQILPSRPNASITGTNG"
4564..5238
/gene="dehH2"
4564..5238
/gene="dehH2"
/function="dehalogenation"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="haloacetate dehalogenase H-2"
/protein_id="BAC81974.1"
/db_xref="GI:34013301"
/translation="MKKIEAIPDMYGTLYDVHVSVDACEKQYPGKGDISVLWRQKQ
LEYAWRLAGQVIKFEATANALTYTCNMKLDCEGSAAMRLTEEYLRLKPFPEVRG
ALRALRQGRMLAILNSGTSTETIHVVHNSGVEGEPEHLISVDSARAYKPHPLAYELG
EEAFGISRESILFVSSNPWDVSGAKAFGYQVCWINRYGFAFDELQGTDFTFVPMDAI
VHLIATV"
/complement(5446..6012)
/gene="orf1"
complement(5446..6012)
/gene="orf1"
/codon_start=1
/transl_table=11
/protein_id="BAC81975.1"
/db_xref="GI:34013302"
/translation="MSEPKIVHEEKAQDFDSQSMYGDYLLHDLMLKAQHPLSPAHD
EMLFIVQHTSELWMKMLHMLMAATRIATEQLPDAPKMLARVSRINEQLVGRVRL
STMTPEYSAMRPYASSSGPSQAYKICIEFALGNKNAAMLKPHARADLLALVDAAI
RMTLSYDEALFLLARGLAVPPGYVERN"
6681..9884
/note="IS1071"
6826..9741
/gene="tnpA"
6826..9741
/gene="tnpA"
/function="transposition"
/codon_start=1
/evidence=experimental
/transl_table=11
/product="transposase"
/protein_id="BAC81976.1"
/db_xref="GI:34013303"
/translation="MQGWHHTFLGMRGLPROISDFEMKAFPTFGAERDAINARRGDS
HKLGLALHIGFLRMSGLLGAFRVIVPALMRHLGNELGIAPEVASLRAMTERGRTLF
DHOQVACTVLGFQWMSHQRSLVRELDEVARCADRDQLLVRARQWLYKKNLVIYVHE
RAIRTLIAAALQLEVTGTATASVDPATLDRWRASVSELRPDGTQQSWLWAPAK

repeat_region
gene
CDS

gene	RKMAENPELASDWDAMREAAARRHELLMRKQEQERKQSQAOERFGRSQSLGSLARPA"
CDS	/complement (847. .1194) /gene="trao" /complement (847. .1194) /gene="trao" /note="ORF2" /codon_start=1 /transl_table=11 /product="Trao" /protein_id="AAR31020.1" /db_xref="GI:39777445" /translation="MKERLLVNGQRIVQAEKDGAWTNQKVDKAGALKPGIYNLYTAQ AADKKQTHAGVIVHADATNVYQIGKNFVHMARSDFDKVPEIGSAKSISYNAQGA AAEAPKLTRGRSM" /complement (1365. .2396) /gene="kfra" /complement (1365. .2396) /gene="kfra" /note="ORF3" /codon_start=1 /transl_table=11 /product="transcriptional regulator" /protein_id="AAR31021.1" /db_xref="GI:39777446"
gene	
CDS	/translation="MAISKEQIFAVADELDAAGNQPTLANVRKOLGSGSFTTISEAMN EMARKASQAPIREPAPQAITDKLAELGDLWVALEMANNRLAAREALEAVRQET EARQEAELADQLTGELDEAKRIAALAEVAAKGEADELRGLAATSRAATAEA RAGELRYELDAHQBQARAEKQACERATASADOVEALRADLAANAASRTAEIERRA GELRADLERNAQDQARAAQAEQRKATEAAERDQVRAELVKVQAKAEAAEQAHQ QRKMAAEHRQAERLTAAQGERDQAKDAQAAREEANGLRLEALETVNAQGRAAD PTSGGRKKT" /complement (2575. .3624) /gene="korB" /complement (2575. .3624) /gene="korB" /note="ORF4" /codon_start=1 /transl_table=11 /product="transcriptional repressor" /protein_id="AAR31022.1" /db_xref="GI:39777447"
gene	
CDS	/translation="MSAKTNAKKEQDKPQSSGLGLDGLAGLNEQPAANAGGAG PQELPLDIDBDHPQRTADNPGFSPESIAIGETIKARGVKSPISVRENPDAGRYL INHGARVYRGSKAHKTTIPAFIDNDYNEADQVIENLQRELTAREIDTIGRELAKG KKGEIAKEIGSLPAFYQHVTLLDPEPIAEAFNSGRKDVTVINELVTAYKKNPB VAAMLADQSLKTRGSKLREFLEDKSHEDGRDPTVDALTKGTDAAEGDGEQGP QDDAKGKKEPADPKLKKAIIVQHDDRPARLLINRRPPAEGNAWLKYEDDGQBF EADLGTVQLVALLBG" /complement (3621. .4703) /gene="incC1" /complement (3621. .4703) /gene="incC1" /note="ORF5" /codon_start=1 /transl_table=11 /product="inclusion membrane protein" /protein_id="AAR31023.1" /db_xref="GI:39777448"
gene	
CDS	/translation="MGVLHESTANRSPIPGSDQGPDRATDHRHSARRAGRWKAPGRV CGLAGDGGGIGAGQPRVGGSGSAPLAGRAGDGTAGASSVHRQALGSRQKQKEET GTQLMKLTIVTAIQGGQKTFATCHLAFDFQERGLRVAVIDLDTQGNASMTLAGHDSG YPASRMFTAGDGLRANFADREDDDLALIAADANLANLDKMLSQAAALRASVAALG EFFDVCLIDTAPSLGVAMTAVALTDVYMLSPIEMEAYSLOQMKQWVAVISNLRKONPK LRFGLWPNKVDARKPHVNNLATLQQAYPQLILPFSIGARDSIAELGEQMPVWKIK KTAARKATQEVRLADYVFTKMEIAQ" /complement (3621. .4385) /gene="incC2" /complement (3621. .4385) /gene="incC2" /note="ORF5.1" /codon_start=1 /transl_table=11 /product="inclusion membrane protein"
gene	
CDS	
gene	/protein_id="AAR31024.1" /db_xref="GI:39777449" /translation="MKTLVTAIQKGGQKTFATCHLAFDFQERGLRVAVIDLDTQGNA SWTLAGHDSGVPASRMFTAGDGLRANFADREDDDLALIAADANLANLDKMLSQAA ALRASVAALGEFFDVCLIDTAPSLGVAMTAVALTDVYMLSPIEMEAYSLOQMKQWVA ISNLRKONPKLRFGLWPNKVDARKPHVNNLATLQQAYPQLILPFSIGARDSIAEL GEQMPVWKIKKTAARKATQEVRLADYVFTKMEIAQ" /complement (4382. .4690) /gene="korA" /complement (4382. .4690) /gene="korA" /note="ORF6" /codon_start=1 /transl_table=11 /product="transcriptional repressor" /protein_id="AAR31025.1" /db_xref="GI:39777450" /translation="MKKRLTEAQFOAAIKGLEICQQTIDIARGVLVDGRPOAEFVASL GLTKGAVSQAVSRVNAVAGEVLPGQFERVTAVLPEHQAFIVKRWADAKGRKQEPNS " /complement (4797. .5126) /gene="kleF" /complement (4797. .5126) /gene="kleF" /note="ORF7" /codon_start=1 /transl_table=11 /product="stable inheritance protein" /protein_id="AAR31026.1" /db_xref="GI:39777451" /translation="MKKQTLPPPGFVEPNTGRVAVLVREYAAASDLNDGADPAYVYSAQ SEEWGLDPWRLVEGVDPHTAGQPDVCFANGSSRTVGPLMTFFMSSAADAARLNAKED HAPIFSR" /complement (5128. .5511) /gene="kleB" /complement (5128. .5511) /gene="kleB" /note="ORF8" /codon_start=1 /transl_table=11 /product="stable inheritance protein" /protein_id="AAR31027.1" /db_xref="GI:39777452" /translation="MLPSGSTQHSSTRKRVQGAQMSNIKFKPKIEQPAAPAPVEPAAA SAPKAPAGAGLAGLVKVFVMAVTLVWPIILKWLSTATFPQFVVRMLYHMTPTGYAG WSFLYPAVLTATYFVSIYKPKGL" /complement (5663. .5899) /gene="kleA" /complement (5663. .5899) /gene="kleA" /note="ORF9" /codon_start=1 /transl_table=11 /product="stable inheritance protein" /protein_id="AAR31028.1" /db_xref="GI:39777453" /translation="MSDNKIMPWIDELEGAATDPPARRDEIAAMMAEAEALVCKABE LRGKAVPAGCSLEQGAQGHWSMEAVEQAKRAGH" /complement (6057. .6314) /gene="korC" /complement (6057. .6314) /gene="korC" /note="ORF10" /codon_start=1 /transl_table=11 /product="transcriptional repressor" /protein_id="AAR31029.1" /db_xref="GI:39777454" /translation="MTNDANIRLECLKPAERWAQPTGEVREVLRLAGFSGSKAKAL GLGAKGDRITRRWIGEDTPIPYAAWAILCDQAGLGVWIKED" /complement (6331. .7539) /gene="kleB" /complement (6331. .7539)

	<pre>/gene="k1cB" /note="ORF11" /codon_start=1 /transl_table=11 /product="stable inheritance protein" /protein_id="AAK31030.1" /db_xref="GI:39777455" DAALMRGDGAAEAASDRYEAVIWKNGGTHVSGSMADHAPGQVIERHCNAPGVDPL</pre>	
Query Match	14.4%;	Score 46.2; DB 1; Length 87688;
Best Local Similarity	47.7%;	Pred. No. 31;
Matches 135; Conservative	0;	Mismatches 148; Indels 0; Gaps 0;
QY	6	CCACAAGCGCACTAAACGCGCAGCAGCATCGCGCAGGCTCAACGCCGCCGCGTCCGAA 65
DB	1848	CGAAGGCCAGAGAGGCGCAGCGCCAGCGCCGACGAGTGGAGGCGCTCGCGGCCGA 1789
QY	66	TCGCGTTGGCGCGCAACATGTTGGCGCGCGAGCTTCCGTCGCGCAGCAGCGCCGCGC 125
DB	1788	CCTCGCGCGCGCAACAGCGCTACCGCGGAGATCGAGCGCGCGCGGAGCTGCGTGC 1729
QY	126	CCAAAGCGCAGCGCACTCAGGCTATCCCGCAGCGTAGCGCGAGATGTCGCGACATC 185
DB	1728	CGACCTGGAGCGGCGCAACAGCGACCGACCGAGCGCGCGCGCTCAGCGCCGAGCAGCG 1669
QY	186	ACACGAATCGACGACGACAGCGCACTGGCGCTGTTGTCGATGCATCTCGACGATGA 245
DB	1668	GAGGCCACCGAGCGCGCGCGCGGACCGGACCGGAGTGGCGCGGAGCTGTGAAGGT 1609
QY	246	GCAGCGCGCGCTTGAGCGCGGATGAAGCTCGCTGCGCATCCG 288
DB	1608	GCAGCGCAAGCGCGCGCGCGCAGCAGCGCCACCGAGGACAG 1566
RESULT 10		
U66917		
LOCUS	108845 bp	DNA circular BCT 03-JUL-2003
DEFINITION	Pseudomonas sp. ADP atrazine catabolic plasmid PADP-1, complete sequence.	
ACCESSION	U66917	AF017572
VERSION	U66917.2	GI:13937422
KEYWORDS	Pseudomonas sp. ADP	
SOURCE	Bacteria; Proteobacteria.	
ORGANISM	Pseudomonas sp. ADP	
REFERENCE	1	(bases 34964 to 36388)
AUTHORS	de Souza,M.L., Wackett,L.P., Boundy-Mills,K.L., Mandelbaum,R.T. and Sadowsky,M.J.	
TITLE	Cloning, characterization, and expression of a gene region from Pseudomonas sp. strain ADP involved in the dechlorination of atrazine	
JOURNAL	Appl. Environ. Microbiol.	61 (9), 3373-3378 (1995)
MEDLINE	96035669	
PUBMED	7574546	
REFERENCE	2	(bases 43552 to 46212)
AUTHORS	Boundy-Mills,K.L., de Souza,M.L., Mandelbaum,R.T., Wackett,L.P. and Sadowsky,M.J.	
TITLE	The atzB gene of Pseudomonas sp. strain ADP encodes the second enzyme of a novel atrazine degradation pathway	
JOURNAL	Appl. Environ. Microbiol.	63 (3), 916-923 (1997)
MEDLINE	97208208	
PUBMED	9055410	
REFERENCE	3	(bases 69775 to 71867)
AUTHORS	Sadowsky,M.J., Tong,Z., de Souza,M. and Wackett,L.P.	
TITLE	AtzC is a new member of the amidohydrolase protein superfamily and is homologous to other atrazine-metabolizing enzymes	
JOURNAL	J. Bacteriol.	180 (1), 152-158 (1998)
MEDLINE	98083068	
PUBMED	9422605	
REFERENCE	4	(bases 43552 to 46212)
AUTHORS	Boundy-Mills,K.L., de Souza,M.L., Mandelbaum,R.T., Wackett,L.P. and Sadowsky,M.J.	
	Direct Submission	
	Submitted (15-AUG-1996) Soil, Water, and Climate, University of Minnesota, 1991 Upper Buford Circle, 439 Borlaug Hall, St. Paul, MN 55108, USA	
	5 (bases 69775 to 71867)	
	Sadowsky,M.J., Tong,Z., de Souza,M.L. and Wackett,L.P.	
	Direct Submission	
	Submitted (05-AUG-1997) Soil, Water, and Climate Dept., University of Minnesota, 1991 Upper Buford Circle, 439 Borh, St. Paul, MN 55108, USA	
	6 (bases 1 to 108845)	
	Martinez,B.M., Tomkins,J., Wackett,L.P., Wing,R. and Sadowsky,M.J.	
	Direct Submission	
	Submitted (23-JAN-2001) Biochemistry, Molecular Biology and Biophysics, University of Minnesota, 1479 Gortner Ave., 140 Gortner Labs, Saint Paul, MN 55108, USA	
	REMARK	
	Nucleotide sequence updated by submitter	
	COMMENT	
	On or before May 3, 2001 this sequence version replaced gi:2736100, gi:1732205.	
	FEATURES	
	Location/Qualifiers	
	1..108845	
	/organism="Pseudomonas sp. ADP"	
	/mol_type="genomic DNA"	
	/strain="ADP"	
	/db_xref="taxon:47660"	
	/plasmid="PADP-1"	
	/note="atraine catabolic plasmid"	
	complement(869..1183)	
	/gene="orf1"	
	complement(869..1183)	
	/gene="orf1"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAK50234.1"	
	/db_xref="GI:13937423"	
	/translation="MSYAVFSPSEAEQALRYIAAASPDVAARYTEAIVTCES LRTFTPLGTMRDVRPGLRITNYRKRVIADFVAGDLVSIIGVFYGGQDYESILHDDA DD"	
	complement(1180..1461)	
	/gene="orf2"	
	complement(1180..1461)	
	/gene="orf2"	
	/codon_start=1	
	/transl_table=11	
	/product="hypothetical protein"	
	/protein_id="AAK50235.1"	
	/db_xref="GI:13937424"	
	/translation="MEAAVRTTQQLSITLPNDMDVVKTKVTKGYASESEVIRDLR ALLARDRAVESWLNQVGPAYDALKADPSRAVTADQVRARLAAEHAQAQ"	
	1618..2046	
	/gene="k1cA"	
	1618..2046	
	/gene="k1cA"	
	/note="protein involved in plasmid stable inheritance"	
	/codon_start=1	
	/transl_table=11	
	/product="k1cA protein"	
	/protein_id="AAK50236.1"	
	/db_xref="GI:13937425"	
	/translation="MNTQDPVTASLVAEQRLDFLPYFGRLMRGALVYAMLR LCERYNGAYWHYYTSLDGGFYLPADLAERLEIVDNGNFRGELSADAAGIVATLFAIG QLAERIAGTDAADALIDRYHFLRGPAAGHPAAAIYRAID"	
	2098..2517	
	/gene="k1cB"	
	2098..2517	
	/gene="k1cB"	
	/note="protein involved in plasmid stable inheritance"	
	/codon_start=1	
	/transl_table=11	
	/product="k1cB protein"	
	/protein_id="AAK50237.1"	
	/db_xref="GI:13937426"	

```
/translation="MAKQTKQAPQAGQDWSAPAEALAEALPADRDLGLAALAAVAAVWEI
DAAYVRGGAAGAAELAGDREYAI IWKLGNTNFGWADDEAAGRVIERRHCAAVPGDVEL
WGQGFQFLAVAGQLARPGVGRULWRAAQRAFSVRG"
3427..3684
/gene="korC"
3427..3684
/gene="korC"
/codon_start=1
/translation="transcriptional repressor"
/translation="KorC protein"
/protein_id="AAKS0238.1"
/db_xref="GI:13937427"
/translation="MTNDANIRLECLPAERWAQPSGSEVREVLRLAGLTGGKAQVL
GLGPKGRTIRRWGEDTPIPYAAWALLCDYAGLGIWKEA"
3844..4080
/gene="klea"
3844..4080
/gene="klea"
/translation="protein involved in plasmid stable inheritance"
/translation="KleA protein"
/protein_id="AAKS0239.1"
/db_xref="GI:13937428"
/translation="MSKNKIMPVDALPNVEATDFQARRDQIEATMAEAAELVKQABE
LRGKAYFAALGLEASAKGEMSSQAVEQAKRSVGW"
4139..4354
/gene="kleB"
4139..4354
/gene="kleB"
/translation="protein involved in plasmid stable inheritance"
/translation="KleB protein"
/protein_id="AAKS0240.1"
/db_xref="GI:13937429"
/translation="MSKGKIEIETCCRRCKGSIRTLSTHTIIGADDAREKFGSICGGC
ITPDEDNELTEMLLAAAVRRMSGATLQ"
4442..4828
/gene="kleG"
4442..4828
/gene="kleG"
/translation="protein involved on plasmid stable inheritance"
/translation="KleG protein"
/protein_id="AAKS0241.1"
/db_xref="GI:13937430"
/translation="MLPSGSIQHSRNRVEQGMNSIVKFPRAKPKPAPEVPQPAAPA
AAPAKAEGRLVAGLVKFWVATLVWPKLVLAITTFQFVRMLYHWNTPGVYA
GWSFLAYPAALTAITYFVSIYKPKGL"
4830..5360
/gene="kleP"
4830..5360
/gene="kleP"
/translation="protein involved in plasmid stable inheritance"
/translation="KleP protein"
/protein_id="AAKS0242.1"
/db_xref="GI:13937431"
/translation="MAKOTLPYPPGFVPTTGRVAVLVREYADSLDNGDAPYVWSAQ
SEEGLDPRWLVGVDPHVGGSDVCFASGTRTVGLMTFELSAAHAAQLIDAKGE
ELAQRLATLAVIADGLGLPAKALRIEAKVEGRPAVFYDQDQATLCACAVDSHDWRQAR
ATAATASAIKARTNF"
5104..5367
/gene="kleQ"
5104..5367
/gene="kleQ"
/translation="protein involved in plasmid stable inheritance"
/translation="KleQ protein"
/protein_id="AAKS0243.1"
/db_xref="GI:13937432"
/translation="MRHSSLTTPRGKSWPCAPPWPSRTGWACLPRRCASRRWRPAGR
PCSTRTYGPAPARAWIPTTGDRGRPPQHRPSTRREQIFSR"
5461..5637
/gene="incC"
5461..5637
/gene="incC"
/translation="inclusion membrane protein"
/translation="InC protein"
/protein_id="AAKS0244.1"
/db_xref="GI:13937433"
/translation="MGAIHEETANRSPIPDGHOGAGDRAADHRHSARRAGRWPAAGGV
CHFAGADGGVAGGQSRMSSRTAPRGLRASDGTAGASSVHRQVGSRRQETGT
QVMTLVTAIQKGGQKTFATCLHAFLERLVRAVIDLDTQGNASFTLSAYQSGYL
```

```
/product="KleG protein"
/protein_id="AAKS0243.1"
/db_xref="GI:13937432"
/translation="MRHSSLTTPRGKSWPCAPPWPSRTGWACLPRRCASRRWRPAGR
PCSTRTYGPAPARAWIPTTGDRGRPPQHRPSTRREQIFSR"
5461..5637
/gene="incC"
5461..5637
/gene="incC"
/translation="inclusion membrane protein"
/translation="InC protein"
/protein_id="AAKS0244.1"
/db_xref="GI:13937433"
/translation="MGAIHEETANRSPIPDGHOGAGDRAADHRHSARRAGRWPAAGGV
CHFAGADGGVAGGQSRMSSRTAPRGLRASDGTAGASSVHRQVGSRRQETGT
QVMTLVTAIQKGGQKTFATCLHAFLERLVRAVIDLDTQGNASFTLSAYQSGYL
```

gene	Qy	6	CCACAAGCGCCTAAACCCAGCCAGCCATCCCGCAGGGCTCAACGCCCGCGTCGGAA	65
CDS	Db	8311	CGACAAGCGCCAGAGAGGCGCACCGCCAGCGCCAGCGTGGAGGCCCTCGCGGCCGA	8370
	Qy	66	TCGCGTTGGCGGCAACATGTTGGCGCGGACGTTCCGTCGCCGAGCAGCGCCGCGC	125
	Db	8371	CCTCGCGCGCGAACACCGCGTACCGCGGAGATCGAGCGCGCGCGCGAGCTGCGTGC	8430
	Qy	126	CCAAAGCGCAGCGGACCTCGAGGCTATCCGCGAGCGGTACCGCGAGATGTTGGCGACATC	185
	Db	8431	CGACCTGGAGCGGCGAACCAGCGACCGCGCGCGCGCGCTCAGCGCGAGCAGCG	8490
	Qy	186	ACAGAAATCGACAGCACAGCGGAACTGGCGCGTGTTCGATGATCTTCGAGATGA	245
	Db	8491	GAAGGCCACCGAGCGCGCGCGCGGACCGTGGCGCGGAGCTGGTGAAGT	8550
	Qy	246	GCAGCGCGCGCTTCAGCGCGGATGAAGCTCGGCTGGCATCCG	288
	Db	8551	GCAGCCAAAGCGCGCGCGCGAGCGCCACACAGAGCAG	8593

Query Match 14.4%; Score 46.2; DB 1; Length 108845;
Best Local Similarity 47.7%; Pred. No. 30;
Matches 135; Conservative 0; Mismatches 148; Indels 0; Gaps 0;

RESULT 11
AE016924/c
LOCUS AE016924 Chromobacterium violaceum ATCC 12472 section 15 of 16 of the complete genome.
DEFINITION AE016924 AE016825
ACCESSION AE016924.1 GI:34105231
VERSION AE016924.1
KEYWORDS
SOURCE Chromobacterium violaceum ATCC 12472
ORGANISM Chromobacterium violaceum ATCC 12472
REFERENCE 1 (bases 1 to 301457)
Bacteria; Proteobacteria; Betaproteobacteria; Neisseriales; Neisseriaceae; Chromobacterium.
AUTHORS Brazilian National Genome Project Consortium
CONSTRM The complete genome sequence of Chromobacterium violaceum reveals remarkable and exploitable bacterial adaptability
TITLE Proc. Natl. Acad. Sci. U.S.A. 100 (20), 11660-11665 (2003)
JOURNAL 2882880
MEDLINE 14500782
PUBMED 2 (bases 1 to 301457)
REFERENCE Vaeconcelos A.T.R., de Almeida, D.F., Almeida, F.C., de Almeida, L.G.P., de Almeida, R., Goncalves, J.A.A., Andrade, E.M., Antonio, R.V., Araujo, M.F.F., Filho, S.A., Azevedo, V., Batista, A.J., Bataus, L.A.M., Batista, J.S., Belo, A., vander Berg, C., Blamey, J., Bogo, M., Bonato, S., Bordignon, J., Brito, C.A., Brocchi, M., Burity, H.A., Camargo, A.A., Cardoso, D.B.P., Carneiro, N.P., Carraro, D.M., Carvalho, C.M.B., Cascardo, J.C.M., Cavada, B.S., Chueire, L.M.O., Pasa, T.B.C., Duran, N., Fagundes, N.,

3010. .3330
 /locus_tag="CV3928"
 3010. .3330
 /locus_tag="CV3928"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA061589.1"
 /db_xref="GI:34105234"
 /translation="MKYNHIGTSGRFDNEIDLPHLKMTVSDHQDNAGFIQWORYWQ
 DAPYELVKKVPHIAFEVDLAQLEGHKLLIAPNSPGLTVAFIEVNGAPVELWQY
 DKNV"
 complement (3435. .3674)
 /gene="rRNA5S"
 /locus_tag="CVrRNA-5s-7"
 complement (3435. .3674)
 /gene="rRNA5S"
 /locus_tag="CVrRNA-5s-7"
 /product="rRNA-5s"
 /note="identified by sequence similarity; putative"
 complement (3675. .6614)
 /gene="rRNA23S"
 /locus_tag="CVrRNA-23s-7"
 complement (3675. .6614)
 /gene="rRNA23S"
 /product="rRNA-23s"
 /note="identified by sequence similarity; putative"
 complement (6813. .6888)
 /locus_tag="CVrRNA-Ala-TGC-7"
 complement (6813. .6888)
 /locus_tag="CVrRNA-Ala-TGC-7"
 /product="tRNA-Ala"
 complement (6975. .7051)
 /locus_tag="CVrRNA-ile-GAT-7"
 complement (6975. .7051)
 /locus_tag="CVrRNA-ile-GAT-7"
 /product="tRNA-ile"
 complement (7295. .8768)
 /gene="rRNA16S"
 /locus_tag="CVrRNA-16s-7"
 complement (7295. .8768)
 /gene="rRNA16S"
 /locus_tag="CVrRNA-16s-7"
 /product="rRNA-16s"
 /note="identified by sequence similarity; putative"
 8828. .9085
 /locus_tag="CV3929"
 8828. .9085
 /locus_tag="CV3929"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA061591.1"
 /db_xref="GI:34105236"
 /translation="MCKYSAPAKHLHLSVHVKQCSRLRRRCQTVSFPLFRSLRQ
 LRRTRDPKPVNTFREKNLRKMPILFQNNKLRKXHL"
 9220. .10809
 /gene="nadB"
 /locus_tag="CV3930"
 9220. .10809
 /gene="nadB"
 /locus_tag="CV3930"
 /FC_number="1.4.3.16"
 /note="identified by sequence similarity; putative; ORF
 located using Glimmer/Genemark/Blastx/COG0029"
 /codon_start=1
 /transl_table=11
 /product="L-aspartate oxidase"
 /protein_id="AA061592.1"
 /db_xref="GI:34105237"
 /translation="MLKFDVLLIGSLAGMTLALHAETRKVGLITKRDLLBGNSTYA
 QGGIAAVLTDSDVEDHIDRTLVAGLGCNEKATRFIVEHSKEAIDWLKIGVFPFTRD

gene
 CDS

gene
 rRNA

gene
 rRNA

gene
 tRNA

gene
 tRNA

gene
 rRNA

gene
 CDS

gene
 CDS

Falcao,C.L., Pantinatti,F., Farias,I.P., Felipe,M.S.S.,
 Ferrari,L.P., Ferro,J.A., Ferro,M.I.T., Franco,G.R.,
 Freitas,N.S.A., Furlan,L.R., Gazzinelli,R.T., Gomes,E.A.,
 Goncalves,P.R., Grangeiro,T.B., Grattapaglia,D., Grisard,E.C.,
 Guimaraes,C.I., Hanna,E.S., Hungria,M., Jardim,S.N., Laurino,J.,
 Leao,L.C.T., Passarella,L., Lima,A., Loureiro,M.F., Lyra,M.C.P.,
 Macedo,M., Madeira,H.M.F., Manfio,G.P., Maranhao,A.Q.,
 Martins,W.C., Mauro,S.M.Z., de Medeiros,S.R.B., Meissner,R.V.,
 Menck,C.F.M., Moreira,M.A.M., Nascimento,F.F., Nicolas,M.F.,
 Oliveira,J.G., Oliveira,S.C., Paixao,R.F.C., Parente,J.A.,
 Pedrosa,F.O., Pena,S.J.D., Pereira,J.O., Pereira,M.,
 Pinto,L.S.R.C., Pinto,L.S., Porto,J.I.R., Potrich,D.P.,
 Neto,C.E.R., Reis,A.M.M., Rigo,L.U., Rondinelli,E., dos
 Santos,E.B.P., Santos,F.R., Schneider,M.P.C., Senanez,H.N.,
 Silva,A.M.R., da Silva,A.L.C., Silva,D.W., Silva,R., Simoes,I.C.,
 Simon,D., Soares,C.M.A., Soares,R.B.A., Souza,E.M., Souza,K.R.L.,
 Souza,R.C., Steffens,M.B.R., Steindel,M., Teixeira,S.R.,
 Urmenyi,T., Vettore,A., Wassem,R., Zaha,A. and Simpson,A.J.G.
 Direct Submission
 Submitted (22-JAN-2003) Labinfo, LNCC - Laboratorio Nacional de
 Computacao Cientifica, Rua Getulio Vargas 333, Petropolis, RJ
 25651070, Brazil

FEATURES
 source

1. .301457
 /organism="Chromobacterium violaceum ATCC 12472"
 /mol_type="Genomic DNA"
 /strain="ATCC 12472"
 /db_xref="taxon:243365"
 121. .1386
 /gene="goaG"
 /locus_tag="CV3926"
 121. .1386
 /gene="goaG"
 /locus_tag="CV3926"
 /EC_number="2.6.1.19"
 /note="identified by sequence similarity; putative; ORF
 located using Glimmer/Genemark/Blastx/COG0160"
 /codon_start=1
 /transl_table=11
 /product="4-aminobutyrate transaminase"
 /protein_id="AA061587.1"
 /db_xref="GI:34105232"
 /translations="MSQQQLMRKSDATPRGVGVMCSFFADRAKNAEVDVTGRRYID
 FAGIGVNLTHLHPKVAEVAEOLSRFSHTCYQVVPYSEYISLAEKLNALAPGKHAK
 KTAFFTPGAENVNAVKIARAATGRPCVIAFGAPGRTLMGALTGKVPAYKLAGFGP
 FPSDVHAPNTLHGVSQDSLGLKLFKFDIDPARVAAIIVPEQGGGFFAAPA
 EWKALRELCDRHGILLIADIEVQSGFARTKLLASHFPAVPLITMAKSLAGGFFLS
 AVTGRAEVDAPAPGLGGTYAGNPLAALAAHVLVDYIDSEKLAERAQRLGDAKLEKL
 NGLKSAVQIAEVRGLGAMIAVFNQPSGHPDPDFAKKVQOHALANGLLILTCGIYA
 NVVRFPLTIEDIEFAELGKLEALKA"
 1456. .2916
 /gene="gabd"
 /locus_tag="CV3927"
 1456. .2916
 /gene="gabd"
 /locus_tag="CV3927"
 /EC_number="1.2.1.16"
 /note="identified by sequence similarity; putative; ORF
 located using Glimmer/Genemark/Blastx/COG1012"
 /codon_start=1
 /transl_table=11
 /product="succinate-semialdehyde dehydrogenase [NAD(P)]"
 /protein_id="AA061588.1"
 /db_xref="GI:34105233"
 /translations="WNLQDPSLLRQCYLGGWLDAGDGHIAVHNPAATGKIGQVP
 KMGAAERAAVAAAPPAWPAKSAKERAAILRWFVDLIIHQDQDLAILTSEOGKP
 LAEAKGEIAYGASTVTAEEAKVIGDIIVPGADRRVLVTKPIGVTAIIPWNPFP
 SAMTRKAPALAGCTVVRPASCTPLSALAEALAEACIAGFAGVSVITGASSELG
 AVLTGSDTVKFTSGFTSGVGRKLIQAQAGTAKVKKVSMELGNAPFIIVDDADLAAVE
 GALTISKYNAQGTVCANRIYVQGVYDAFAKEFAAAVATLKVNGLEAGVSGQLFID
 HNAVKVEHIALADAKGRRIAGKRGHALGHTFEPTLLIADVTRDMKVARIEETFGPL
 ALPFRFKTEEARILANDTFGLASIFYVARDVGRIFRVAEGLEYGMVAVNSGLLSNEA
 APFGCVKOSGLGREGSKYGIEDYLEIKVLLGLDR"

gene
 CDS

DEHOTGHLTREGGSHRIIHAADATGAATVSTLQKIKAPNITELEHYAIDLIT
SOKLQSKRAYGVYALDKNQDEVYTAOHTILASGAGKGVLYTNPDTATCGDIA
MGWAGCQVNMERTOEPTCTLYHPHAKSFLISEAVRGEGLKLPDGTGEMPOHDP
GELAPDVVARAIEPHEKKGDLCDVYLDISYKPAQFIKEHPNFIYACLEGLDITQO
PIPVVAFHTCGGIVTLAGKTDVGLIYAGEVACTGLHGANKRANLSLLECLVIGK
AAAEIDLAADSIKILPAPDWDSDRVTDPDEEVIWISHNMDLRAMWDYGVIVTKRL
YPRLLIDLLKEIDEYSHFHSVNDLIELNLVQTAEILVRSALLRKESRGLHSRD
YPRLLPEAAETVLAADPLGG"
11124. 12206
/locus_tag="CV3931"
11124. 12206
/locus_tag="CV3931"
/EC number="3.2.1.132"
/note="identified by sequence similarity; putative; ORF
located using GeneMark/Blastx"
/codon_start=1
/transl_table=11
/product="probable chitinase A"
/protein_id="AAQ61593.1"
/db_xref="GI:34105238"
/translation="MKYAAKRLSPSATLGLALLAAGSAAAGTAGSPEROTVRL
PDCAANSASPYPQAGSVASRGVNYTAFTQGNPPEQGOAGKRCRPAQAQAA
DHDANFPATLKFYKANTGLGEOQNDIMKLNKPEQDSLEWTFYGCENIGKRGY
TIGIFGATVGGNDEGPPDPLFEKFDASSGASNPISAGGLARAGVHSGMGKILKIS
DSAKVFCDKIGSLQNPAMDAMNTFYKYIYQVQOARQGFSSALTIGSFYDITL
NQGATGSDGTLQGLLSRSGSGDEKTFMTAFYQSRKIVTDNDYQNPNGKRVKQWS
TLNNGETDLKNADAQVKVTWEMK"

Query Match 14.1%; Score 45.4; DB 1; Length 301457;
Best Local Similarity 51.8%; Pred. No. 35;
Matches 103; Conservative 0; Mismatches 96; Indels 0; Gaps 0;

QY 104 CGTCCGCGAGCAGCGCGCCCAACGCGGCGGACCTCGAGGCTATCCGCGAGCGT 163
DB 170537 CTTGGCGGAAACCGGCCAACGAGGCGCGCTCGCGCGCGCTGCGCCACCGAGCTGGCGG 170478

QY 164 AGCGCGAGATGTGCGGCATCACAGAAATCGACGACGACACACGACCGAATCGCGCTGT 223
DB 170477 GCCTCGACGCGCGGATCCAGGCGGAAGTGAAGACGAAACGCGCTGGGAGCGCCAGC 170418

QY 224 TGTGATGATCTCGACATGAGCGCGCGCTTGAGCGCGGAGTGAAGCTCGCTGCG 283
DB 170417 TGTGGAAGATGCCAGCACACACGCGCTGCGCGGAGCTGCTTCATGCGCGCGCTGCG 170358

QY 284 ATCCGTATCACCTTCCCGCA 302
DB 170357 GCATATATTCGCCCGCA 170339

RESULT 12
AP005033 320150 bp DNA linear BCT 10-MAY-2003
LOCUS Streptomyces avermitilis genomic DNA, complete genome, section 13/30.
DEFINITION AP005033 BA000030
ACCESSION AP005033.1 GI:29606546
VERSION AP005033.1 GI:29606546
KEYWORDS Streptomyces avermitilis MA-4680
SOURCE Streptomyces avermitilis MA-4680
ORGANISM Streptomyces avermitilis MA-4680
Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.

REFERENCE 1
AUTHORS Omura.S., Ikeda.H., Ishikawa.J., Hanamoto.A., Takahashi.C., Shinose.M., Takahashi.Y., Horikawa.H., Nakazawa.H., Oonoe.T., Kikuchi.H., Shiba.T., Sakaki.Y. and Hattori.M.
Genome sequence of an industrial microorganism Streptomyces avermitilis deducing the ability of producing secondary metabolites
Proc. Natl. Acad. Sci. U.S.A. 98 (21), 12215-12220 (2001)
21477403
MEDLINE 11572948
PUBMED
REFERENCE 2
AUTHORS Ikeda.H., Ishikawa.J., Hanamoto.A., Shinose.M., Kikuchi.H.,

Shiba.T., Sakaki.Y., Hattori.M. and Omura.S.
Complete genome sequence and comparative analysis of the industrial microorganism Streptomyces avermitilis
Nat. Biotechnol. 21 (5), 526-531 (2003)
22608306
PUBMED
REFERENCE 3 (bases 1 to 320150)
AUTHORS Omura.S., Ikeda.H., Ishikawa.J., Hanamoto.A., Takahashi.C., Shinose.M., Takahashi.Y., Horikawa.H., Nakazawa.H., Oonoe.T., Kishida.N., Director-General of Biotechnology Center, Shiba.T., Sakaki.Y. and Hattori.M.
Direct Submission
Submitted (29-MAR-2002) Director-General of Biotechnology Center, National Institute of Technology and Evaluation, Biotechnology Center; 2-49-10 Nishihara, Shibuya-Ku, Tokyo 151-0066, Japan
(E-mail:bioinite.go.jp, URL:http://www.bio.nite.go.jp/, Tel:81-3-3481-1933, Fax:81-3-3481-8424)
This work was done in collaboration with Haruo Ikeda(*1), Jun Ishikawa(*2), Akiharu Hanamoto(*3), Chigusa Takahashi(*3), Mayumi Shinose(*3), Hiroshi Horikawa(*4), Hidekazu Nakazawa(*4), Tomomi Oonoe(*4), Yoshihiro Kishida(*4), Hisashi Kikuchi(*4), Tadayoshi Shiba(*5), Yoshiyuki Sakaki(*6,*7), Masahira Hattori(*1,*7) and Satoshi Omura(*1,*3).
Final finishing process and all annotation were done by H. Ikeda and J. Ishikawa.
*1 Kitasato Institute for Life Sciences, Kitasato University
*2 National Institute of Infectious Diseases
*3 The Kitasato Institute
*4 National Institute of Technology and Evaluation
*5 School of Science, Kitasato University
*6 Institute of Medical Science, University of Tokyo
*7 RIKEN, Genomic Sciences Center
Following url is also available.
http://avermitilis.ls.kitasato-u.ac.jp.
Location/Qualifiers
1. 320150
/organism="Streptomyces avermitilis MA-4680"
/mol_type="genomic DNA"
/strain="MA-4680"
/db_xref="taxon:227882"
/note="This strain is also named as strain: ATCC 31267, NCIMB 12804 or NRRL 8165.-synonym: Streptomyces avermectinus"
complement(134. 11953)
/gene="olmA3"
complement(134. 11953)
/gene="olmA3"
/note="SAV2897
oligomycin gene cluster"
/codon_start=1
/transl_table=11
/product="modular polyketide synthase"
/protein_id="BAC70608.1"
/db_xref="GI:29606547"
/translation="MANEQQLRDLKRVASDLHTRORLTFIESQGTPIAVGNACR YPGVSSPEDILMRVLTDTDAIGPPTDRGWDLDGLHPDPDHPGTCYAEAGGFVEDI ASFDAFGISPREAQMSDPOORLLETWEALHAGLDIHALRGSRGTGVPAGLNQD YGTLAAAPDGLSDYSGTNSVLSGRISYVGLGPAVTVDTACSSSLVTLHAAQ ALRSGCDLALAGATLTLSAVHVALSGORALAPDGRSKAFSAADGAGHSGVGL AVERSDARRIGHRVLAIVGSAVNDQASNGLTAPNGFSQQRVIRQLANAGVSPAE VDVEARHTGSLGDIADALLATYQAKTAPRLMIGSLKSNIGHSGAAGVGVI KMQALRHGLLPRTLHAEPTPNVDSGSAVELLTPPLPWSLPSGAPALRAQARALEH TNHAVILEQAPETPEAPEPTGCEATGATPPLPWSLPSGAPALRAQARALEH LARPGSADVALSLAVTRTGLEHRAVLAARDGFLDALDADRPAEGLRSGPP VDGKVVFFPQCGQWAGMARELLDTSVFAAKAACAAAEFEFVDKVLVLRKPG AASMERIEVQPVLTVMVSLAEWLRSYGVFSDAVSSQSGEIAAAHVAGGLTQDA RVICRISALLAETLVGKVASVADPVERLTRWADRLSLVAGNPRIVAVAGDI DALAEFTAECARDNVRARIVAATVPTHCLVDPRLERLLDLAPVAPRTVPLVSTV TGRDLTAAMDAGVYDNTRAPLPEPVYRALLAEHSAFVSSHPPLAMGVEQTV AAGASGVVVSIESRDEGLGRLMTASLAKHGGVYVDVMPAFVAGTAVLTPTAFQ RTRYWTEAGDRTGDSVGLSDVHPLLCALYRMADGDAVLTGRLSRAQTWLAGHG VSDRVIFPGTGFVELAVLQDQVCGGRIEELTLPLVVPESGAVVQAVQAEADETG ARASVVSRRADDSADTWTWRHSGVLAEPGEPEGITGTWPPEGASVDISGMYDR

Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

Comparison of the genomes of two Xanthomonas pathogens with differing host specificities
 Nature 417 (6887), 459-463 (2002)
 22022145
 12024217
 2 (bases 1 to 12201)
 da Silva, A.C.R., Ferro, J.A., Reinach, F.C., Farah, C.S., Furlan, L.R., Queglio, R.B., Monteiro-Vitorello, C.B., Van Sluys, M.A., Almeida Jr., N.F., Alves, L.M.C., do Amaral, A.M., Bertolini, M.C., Canargo, L.E.A., Camarotte, G., Cannavan, F., Cardozo, J., Chamberg, F., Chapina, L.P., Cicarelli, R.M.B., Coutinho, L.L., Cursino-Santos, J.R., El-Dorry, H., Faria, J.B., Ferreira, A.J.S., Ferreira, R.C.C., Ferro, M.I.T., Formighieri, E.F., Franco, M.C., Greggio, C.C., Gruber, A., Katsuyama, A.M., Kishi, L.T., Leite Jr., R.P., Lemos, E.G.M., Lemos, M.V.F., Locali, E.C., Machado, M.A., Madeira, A.M.B.N., Martinez-Rossi, N.M., Martins, E.C., Meidanis, J., Menck, C.F.M., Miyaki, C.Y., Moon, D.H., Moreira, L.M., Novo, M.T.M., Okura, V.K., Oliveira, M.C., Oliveira, V.R., Pereira Jr., H.A., Rossi, A., Sena, J.A.D., Silva, C., de Souza, R.F., Spinola, L.A.F., Takita, M.A., Tamura, R.E., Teixeira, E.C., Tezza, R.I.D., Trindade dos Santos, M., Truffi, D., Tsai, S.M., White, F.F., Setubal, J.C. and Kitajima, J.P.

Direct Submission
 Submitted (28-NOV-2001) Departamento de Bioquímica, Universidade de São Paulo, Av. Prof. Lineu Prestes 748, São Paulo, SP 05508-900, Brazil

FEATURES

Location/Qualifiers
 1..12201
 /organism="Xanthomonas campestris pv. campestris str. ATCC 33913"
 /mol_type="genomic DNA"
 /strain="ATCC 33913"
 /db_xref="taxon:33913"
 /note="pathovar: campestris"
 68..1018
 /gene="XCC1538"
 68..1018
 /gene="XCC1538"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA040833.1"
 /db_xref="GI:21112614"
 /translation="MRWOLRGAGVGVILVWACSGTSGSGAAEPSSARADRGYAAALHARRAQLHAPDGLSYTSGVRVCRVGVADPHADIVLPAGPAQLGNLTLPDGLRQAAAGTGVTLHGKPVTEVLAPERAGQPGRLRIDAREFVVGQALYGMWRPAALAQGELOYFFPDHWRIDAQWQPDAPRATLLTSTGTPVQVPGAEVFTNRQYRLRPQAQGDARLFPVFADRTSGKESGGARYLYTDPADRGVVLDENLAENPPCALTPVVCPTAPPDNRDLAVNAGEKTYRPSGR"
 complement (1028..1603)
 /gene="XCC1539"
 complement (1028..1603)
 /gene="XCC1539"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA040834.1"
 /db_xref="GI:21112615"

/translation="MSDAVLPSPPITVAPLSYESRLARTLRNIMVVIHCTELPDMA MARYGERVLYDSGTSGNHYIDRNGSIQOYVALERVAHVHVRGHNPHTLGIELVNSG RYHMLDSHQOOREPYDQAINVLLQLEQLPSVRRMAGHEALDTTQEAASDD PSRKIQKRDPGGLPFWHERIETAVRWVHAP"
 1627..22118
 /gene="XCC1540"
 1627..22118
 /gene="XCC1540"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA040835.1"
 /db_xref="GI:21112616"
 /translation="MALNPDSPLGKMATLPQTGQVTVGVPRPARDVAMLEVDAAQAT TGCLGVDRYKSGSGKRGITLLQAEHLPTVIALAGQAALAPATLRNLYVGVILIAL KGRFRIGEVELEGTDCPCSRMEALPGGYNAMRGHGLCARILKGGVIRIGDTV QAL"
 2115..2642
 /gene="XCC1541"
 2115..2642
 /gene="XCC1541"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA040836.1"
 /db_xref="GI:21112617"
 /translation="MSRGHCILAHFSGSPDALKVTALEVAERLGMWTHERPDPFDLMD ARDLGQGDVGRGLRLLEIARATKGPVVLASGSLGSYIAAQVSLQVPTRALFLM VPPTKMGPLPALDAAAVPISIVAHWDELIPADVIWAQARSARLLIVDDGHRLLGAH VQARSARFAELQSL"
 2687..3193
 /gene="XCC1542"
 2687..3193
 /gene="XCC1542"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="acetyltransferase"
 /protein_id="AA040837.1"
 /db_xref="GI:21112618"
 /translation="MSVLIREASDGAALIEVLTMVAFMRAEHSRHSQHVIAALRND DALSLVADHDGYVGHGLAVSPVALSDGSGWFGALGPLAVPGHQGLGTLRLVQAA LATURERGAAGCAVHGSGPFRRLGFAVEPGLSVPDVPASELQALAFGDRLLPLADVA YHPAFGLG"
 3327
 /gene="XCC1543"
 3327
 /gene="XCC1543"
 /note="putative; ORF located using Glimmer/Genemark"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA040838.1"
 /db_xref="GI:21112619"
 /translation="MQHDDQPLPQTAPHGRVISHLGETDQATLLASIKTRLRHND RSAAMEHLQLQVDALWAMELGRFLNRNLGNWHTQVLYAGTLVQQTTPALEYE LFTPTVALTRFRFGIFRTQALRPGMTLASLPCGMADVLSLDYGCQCPDVRLLG IDLPDGLDGLARLAQVQGLEQLTQREDAPGPNPSVDLSNGLNLYEPDDARY TALVRFDRLRPGGTIVASFLTPPPLCAESPWKLDLIDGAALVLQSQVFDVVQAR WASFRTHAQTRETLQESAGTDRVFRIDDRASIFPTVIARKPMSD"
 4693..6306
 /gene="XCC1544"
 4693..6306
 /gene="XCC1544"
 /note="identified by sequence similarity; putative; ORF located using Blastx/Glimmer/Genemark"
 /codon_start=1
 /transl_table=11

/product="ABC transporter ATP-binding protein"
/protein_id="AA040839.1"
/db_xref="GI:21112620"
/translation="MCPSPITLESVYLLPDGSPLESDLTTHDQRTGLVGRNGVVK
SVLRILAGALFAGRCVGTGSHVLPQVSAAPGOSTAALAGVAPLLDLDRIANG
SVAADPFVGRWTHQTFVALEQHGGLGRDALADPTLSGGEAMRVLAGAWLAQ
PDFLLIDEPNHLIDVQRLARLLAQAWTGLLIVSHDRALLETHQHILEPQGLRS
YGNYSAYAAORALEROACDLASRWKRGEOALRCQOQROQROGRASQA
NOAPILLGRKORSEASGLHQHSAERARLATAVIDALQVDAQAVALPPSPA
TAARLATQVVALAGLGRVLDLLHGQPRIGVVGNGSKSLLLGLAQGLLPS
AGCAIHVPVLDQDLQPARSAOHLHAAAPDQQAORLARLRLAQOARADL
PTGLSGGKRLKVALGALYRPAQLLLIDEPNHLIDLASVEAVALLQDVRGALMV
VSHDAAPLQRLGQRLDQAQWRLAPW"
/gene="XCC1545"
/db_xref="XCC1545"
/db_xref="XCC1545"
/notes="identified by sequence similarity; putative; ORF
located using Blastx/Glimmer/Genemark"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AA040840.1"
/db_xref="GI:21112621"
/translation="MPAPMADHLLPPAAATAALHRRHPAVLLPRSPPAARSSPSIPA
YVFFASCAQGLYLLADELLALGASKATATSGNVVGGALRDAQRAVLWSLGRVL
WPLTERDCPDALYAGVAELPWEHLSTGHTLSVDAHVSGTATTHARVAAORIKDAV
VDTRRQGLRSPVDSVDPDLRLNLSLRGRATISVDLGGPLHRGWRMAQNEAPLK
ENLAVALLRAGPRAYADGGLLDPMCSGTLLEGALMAADVAPGLRGYSDIPSR
WRGFRDSDQOLVTEAREDSVGRALKQVINGSDMDPAIRAAENQAVAGVABAIW
FGVREVDLQRTQVGVVPPNPPYDERLAADAALYRLKGLDLPORVPMASLLCGN
AELAYATGLRAGKYQLFNAGIECALIVCDPIAVPRTPLAAPTALSGAQMVAANLR
KNLQFKKWRAREGICFRVYDADLPEYSAADIVYQADGDRRIPLHVOEYAAPATIP
EADVRRLGELHAAAREVEPAERVALKSRERGKSGKYGFEOERNEIVNREHGL
LRVNLFDYLDLTLFLDRPLRGTMAQSKGRFLNLCYTGTVASVQAAVAGASATSV
DLSGTQLQCADMLNAGSGSKHUVQADALAWLEAARHFDVIFCDPTFSNSARA
EDPDIQREHVRLLRAAVARLAPGGVLYFNFRFKLDDEAVSEFAQCEIISPTIDP

Query Match 13.8%; Score 44.4; DB 1; Length 12201;
Best Local Similarity 48.8%; Pred. No. 93;
Matches 120; Conservative 0; Mismatches 126; Indels 0; Gaps 0;
QY 22 CGCAGCAGCCATCGCGCAGGCTCAACGCCCGCGTGGGAATCGGTCGGCGCAA 81
DB 7248 CTCAAGCAGGTGATCCATGACGAGCATGACCCCGCATGCGATTCGCGCGCAAGGAA 7307
QY 82 CATGTTGCGCGCGCAGCTTCGTCGCGGAGCAGCGCGCCCAACGCGAGCGGAC 141
DB 7308 AATGCCAGTGGCGCGCGGAGGCGCATCTGTTGCGCGTGGCGAGGTTCGCGAC 7367
QY 142 CTCGAGGCTATCGCGCAGCGTACGCGAGATGGTGGCGACATCAACGAAATCGACGAC 201
DB 7368 CTGACAGCTCGCGCAGGCGCACCGCGGTGGTGTGCAACCGCGGTACGAGCGG 7427
QY 202 GACACAGCGAACTGGCGCTGTTGTGATGATCATCTCGAGATGAGCAGCGCGGCTTGAG 261
DB 7428 CTGCGCGCGCAGCGCGCTTGTATCGCAAACTTGGCGACACCCCTGACGCGCGTGGCG 7487
QY 262 GCGGGG 267
DB 7488 CAGTGG 7493

RESULT 14
CNS08C7V/c 104289 bp DNA linear PLN 22-JUL-2004
LOCUS
DEFINITION
Oryza sativa chromosome 12, . BAC OSUNBa0039K19 of library OSUNBa
from chromosome 12 of cultivar Nipponbare of sep. japonica of Oryza
sativa (rice), complete sequence.
AL731746
AL731746.5 GI:50470847
KEYWORDS
HTG.
Oryza sativa (japonica cultivar-group)

ORGANISM
Oryza sativa (japonica cultivar-group)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Ehrhartoideae; Oryzaceae; Oryza.
REFERENCE
1 (bases 1 to 104289)
AUTHORS
Choisne, N., Orjeda, G., Cattolico, L., Demange, N., Wincker, P.,
Segurens, B., Pelletier, E., Scarpetti, C., Salanoubat, M.,
Weissenbach, J. and Quetier, F.
Oryza sativa chromosome 12 sequencing
Unpublished
2 (bases 1 to 104289)
REFERENCE
Direct Submission
Genoscope.
TITLE
Submitted (21-JUL-2004) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT
On Jul 21, 2004 this sequence version replaced gi:50399590.
Center: Genoscope / Centre National de Sequencage
Center code: GS
Web site: http://www.genoscope.cns.fr/
Contact: SeqRef@genoscope.cns.fr

The following sequence is oriented from the T7 to the SP6 end.
Upstream BAC (overlapping the T7 end) : OSUNBa0018L16 (AC=BX255875)
Downstream BAC (overlapping the SP6 end) : OJ1536_D01 (AC=AL731752)

FINISHED SEGMENT STARTS AT BASE 1
FINISHED SEGMENT ENDS AT BASE 104289

FEATURES
source
Location/Qualifiers
1..104289
/organism="Oryza sativa (japonica cultivar-group)"
/mol_type="genomic DNA"
/cultivar="Nipponbare"
/sub_species="japonica"
/db_xref="taxon:39947"
/chromosome="12"
/clone_lib="OSUNBa"
/note="group: japonica cultivar-group"

Query Match 13.8%; Score 44.2; DB 8; Length 104289;
Best Local Similarity 50.2%; Pred. No. 70;
Matches 109; Conservative 0; Mismatches 108; Indels 0; Gaps 0;
QY 48 CAAACCGCGCGTCCGAATCGGTTGGCGGCAACATGTTGGCGCGGAGCTTCGCTC 107
DB 60680 CGACGCGCACGCCGACGCCGCGCAAGCGCGCGCACGCCACGCGCGCTTCGC 60621
QY 108 CGCCGAGCAGCGCGCGCGCCCAACGCGCAGCGACCTCGAGGCTATCCGCGGAGCGTACGC 167
DB 60620 CGCCGCGCGGCAATGCGCGCGCGCGCGCACGCCGCGCGCGCTTCGCGGCGGC 60561
QY 168 CGAGATGTTGGCGCATCACACGAAATCGACGACACACAGCCGAACTGGCGCTTTGTC 227
DB 60560 GCGCATGTTGCGCGCGCGCACGCCACGCGCGCGCTTCGCGCGCGCGCAATGCCGCG 60501
QY 228 GATGATCTCGACGATGAGCAGCGCGCGCTTGAGCG 264
DB 60500 GCGCGCACGCCGCGCGCGCGCGCGCGCGCGCGCACGCCACG 60464

RESULT 15
LMFLCHR34_10/c
WPCOMMENT
Sequence split into 18 fragments LOCUS LMFLCHR34 Accession AL499623
Fragment Name Begin End
LMFLCHR34_00 1 110000
LMFLCHR34_01 100001 210000
LMFLCHR34_02 200001 310000
LMFLCHR34_03 300001 410000
LMFLCHR34_04 400001 510000

LMFLCHR34_05 500001 610000
LMFLCHR34_06 600001 710000
LMFLCHR34_07 700001 810000
LMFLCHR34_08 800001 910000
LMFLCHR34_09 900001 1010000
LMFLCHR34_10 1000001 1110000
LMFLCHR34_11 1100001 1210000
LMFLCHR34_12 1200001 1310000
LMFLCHR34_13 1300001 1410000
LMFLCHR34_14 1400001 1510000
LMFLCHR34_15 1500001 1610000
LMFLCHR34_16 1600001 1710000
LMFLCHR34_17 1700001 1720777

Continuation (11 of 18) of LMFLCHR34 from base 1000001 (AL499623 Leishmania major chromo

Query Match 13.8%; Score 44.2; DB 2; Length 110000;
Best Local Similarity 54.7%; Pred. No. 69;
Matches 88; Conservative 0; Mismatches 73; Indels 0; Gaps 0;
Qy 112 GAGCAGCGCGCGCCCAACGGCAGCGGACCTCGAGGCTATCCGCCGAGCGTACGCCGAG 171
Db 15456 GAGCAGCGCGGATGCACAGCTGGAGGGCGTGTGCTGCTGCCATGGGCTTGAGCGGTCGCG 15397
Qy 172 ATGGTGGGACATCACAGAAATCGACGACACAGCCGAACTGGCGCTGTGTGCGATG 231
Db 15396 ATGGAGCTCATCTCGAACGAGATCGACGCCGAGTTGCAGGGGAGAGATCAAGTCATCAAG 15337
Qy 232 CATCTCGACGATGACGAGCGCGGCTTGAGGGGGATGAA 272
Db 15336 GACCTGCACCTCGAGATGCGCGGACTGGAGAGAAGGTGAA 15296

Search completed: February 10, 2005, 06:08:07
Job time : 2027 secs

This page Blank (usps)